

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 6e-20
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 1e-15
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 2e-14
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-09
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YKL179c] 3e-09
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 2e-07
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-07
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c] 2e-07
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 1e-06
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c] 3e-06
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YDR217c] 4e-06
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 2e-05
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YAL035w] 2e-04
[FUNCAT] r general function prediction [M. jannaschii, MJ1254] 0.001
[BLOCKS] BL00387A
[BLOCKS] BL00411H
[BLOCKS] BL00411G
[BLOCKS] BL00411F
[BLOCKS] BL00411E Kinesin motor domain proteins
[BLOCKS] BL00411D Kinesin motor domain proteins
[BLOCKS] BL00411C Kinesin motor domain proteins
[BLOCKS] BL00411B Kinesin motor domain proteins
[BLOCKS] BL00411A Kinesin motor domain proteins
[SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 2e-68
[SCOP] d2tmab 1.105.4.1.1 Tropomyosin [rabbit (Oryctolagus cuniculus) 4e-05
[SCOP] d3kar 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyces 2e-09
[EC] 3.6.1.32 Myosin ATPase 5e-25
[PIRKW] nucleus 4e-27
[PIRKW] phosphotransferase 3e-16
[PIRKW] duplication 6e-20
[PIRKW] citrulline 6e-18
[PIRKW] tandem repeat 4e-24
[PIRKW] heterodimer 3e-28
[PIRKW] endocytosis 1e-23
[PIRKW] heart 1e-17
[PIRKW] transmembrane protein 2e-28
[PIRKW] serine/threonine-specific protein kinase 3e-16
[PIRKW] zinc finger 1e-23
[PIRKW] surface antigen 2e-16
[PIRKW] DNA binding 1e-25
[PIRKW] metal binding 1e-23
[PIRKW] muscle contraction 4e-24
[PIRKW] heterotetramer 4e-24
[PIRKW] acetylated amino end 2e-19
[PIRKW] actin binding 5e-25
[PIRKW] mitosis 3e-58
[PIRKW] microtubule binding 3e-58
[PIRKW] ATP 3e-58
[PIRKW] thick filament 4e-24
[PIRKW] phosphoprotein 9e-29
[PIRKW] leucine zipper 1e-12
[PIRKW] skeletal muscle 8e-24
[PIRKW] disulfide bond 1e-12
[PIRKW] heterotrimer 1e-29
[PIRKW] calcium binding 6e-18
[PIRKW] alternative splicing 4e-21
[PIRKW] P-loop 2e-63
[PIRKW] coiled coil 3e-58
[PIRKW] heptad repeat 1e-25
[PIRKW] methylated amino acid 4e-24
[PIRKW] peripheral membrane protein 1e-23
[PIRKW] dimer 1e-12
[PIRKW] cardiac muscle 1e-17
[PIRKW] hydrolase 5e-25
[PIRKW] microtubule 6e-15
[PIRKW] muscle 7e-23
[PIRKW] membrane protein 6e-20
[PIRKW] GTP binding 8e-22
[PIRKW] EF hand 6e-18
[PIRKW] cell division 1e-25
[PIRKW] cytoskeleton 4e-24
[PIRKW] hair 6e-18
[PIRKW] Golgi apparatus 8e-24
[PIRKW] calmodulin binding 1e-23

[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 3e-16
 [SUPFAM] myosin motor domain homology 5e-25
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-13
 [SUPFAM] kinesin-related protein KIF1 9e-27
 [SUPFAM] kinesin-related protein CIN8 4e-36
 [SUPFAM] kinesin heavy chain 4e-24
 [SUPFAM] plectin 1e-13
 [SUPFAM] trichohyalin 6e-18
 [SUPFAM] kinesin-related protein KIF3 1e-29
 [SUPFAM] kinesin-related protein KIF2 3e-20
 [SUPFAM] ribosomal protein S10 homology 1e-13
 [SUPFAM] giantin 8e-24
 [SUPFAM] protein kinase homology 3e-16
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-13
 [SUPFAM] kinesin-related protein unc-104 8e-26
 [SUPFAM] human early endosome antigen 1 1e-23
 [SUPFAM] unassigned kinesin-related proteins 1e-28
 [SUPFAM] Mycoplasma genitalium hypothetical protein MG218 4e-17
 [SUPFAM] myosin heavy chain 5e-25
 [SUPFAM] conserved hypothetical P115 protein 4e-20
 [SUPFAM] centromere protein E 5e-24
 [SUPFAM] calmodulin repeat homology 6e-18
 [SUPFAM] kinesin-related protein KLP61F 1e-25
 [SUPFAM] hypothetical protein MJ0914 3e-12
 [SUPFAM] kinesin-related protein MKLP-1 2e-63
 [SUPFAM] pleckstrin repeat homology 8e-26
 [SUPFAM] hypothetical protein MJ1322 4e-13
 [SUPFAM] kinesin-related protein KIF1B 3e-28
 [SUPFAM] kinesin motor domain homology 2e-63
 [SUPFAM] kinesin-related protein KLPA 7e-25
 [SUPFAM] kinesin-related protein nodA 1e-12
 [SUPFAM] kinesin-related protein Eg5 5e-30
 [PROSITE] ATP_GTP_A 1
 [PFAM] Kinesin motor domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW COMPLEXITY 7.53 %
 [KW] COILED_COIL 19.78 %

SEQ MESNFNQEGVPRPSYVFSADPIARPSEINFDGIKLDLSHEFSLVAPNTEANSFESKDYLO
 SEG
 COILS
 3kar-

SEQ VCLRIRPFTQSEKELESEGCVHILDSQTIVLKEPQCILGRLSEKSSGQMAQKFSFSKVFG
 SEG
 COILS
 3kar-

SEQ PATTQKEFFQGCIMQPVKDLLKGQSRIFTYGLTNSGKTYTFQGTENIGILPRTLNVLF
 SEG
 COILS
 3kar-

SEQ DSLQERLYTKMNLKPHRSREYLRSLSEQEKEEIASKSALLRQIKEVTVHNDSDDTLYGSL
 SEG
 COILS
 3kar-

SEQ TNSLNISEFEESIKDYEQANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKML
 SEG
 COILS
 3kar-EEEEEEEEETTEEEETTTCC-----CCEE

SEQ RLSQDVKGYSFIKDLQWIQVSDSKEAYRLKLGKIQSVAFTRKLNASSRSHSIFTVKIL
 SEG
 COILS
 3kar- EETTTTE-EEETTCCEEECCGGHHHHHHHHHHHCCTTTTCHHHHHHCEEEEEEEEE

SEQ QIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGERLRETGNINTSLLTLGKCINVLKNS
 SEG
 COILS
 3kar- E--EETTTTCEEEEEEEEEEECCCCC---CCCHHHHHHHHHHHHHHHHHHHHHHTT

SEQ ESKSFQQHVPFRESKLTHYFQSFFNGKGKICMIVNISQCYLAYDETLNVLFKFSATAQKVC
 SEG
 COILS
 3kar- TTTT--TCCTTTTTHHHHHHGGGCTTTTEEEEEEECCGGHHHHHHHHHHHH.....

SEQ VPDTLNSSQDKLFGPVKSSQDVSLDSNSNSKILNVKRATISWENSLEDLMEDEDLVEELE

```

SEG .....XXXXXXXXXXXXXXXXXXXX
COILS .....
3kar- .....

SEG NAEETQNVETKLLDEDLTKTEENKAFISHEEKRKLLDLIEDLKKKLINEKKEKLTLEFK
SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG IREEVTQEFQYWAQREADFKETLLQEREILEENAERRLAIFKDLVGKCDTREEAAKDIC
SEG .....
COILS CCCCCC.....
3kar- .....

SEG ATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRENESDSLIOELETSSNKKII
SEG .....
COILS .....CCCCCCCCCCCCCCCC
3kar- .....

SEG TQNQRIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSSLIINNKLICNETVEVP
SEG .....
COILS CCCCCCCCCCCCCC.....
3kar- .....

SEG KDSKSKICSERKRVNENELQQDEPPAKKGSIHVSSAITEDQKKSEEVPRNIAETIEDIRVL
SEG .....
COILS .....CCCC
3kar- .....

SEG QENNEGLRAFLITIENELKNEKEEKAELNKQIVHFQQLSLSEKKNLTLSKEVQQIQSNY
SEG .....XXXXXXXXXXXXXXXXXXXX
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
3kar- .....

SEG DIAIAELHVQKSKNQEQEEKIMKLSNEIETATRSITNNVSIKLMHTKIDELRTLDSVSQ
SEG .....
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG ISNIDLLNLRDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAI
SEG .....
COILS .....
3kar- .....

SEG WEECKEIVKASSKKSHQIEELEQQIEKLQAEVKGKIDENNRLKEKEHKNQDDLLKEKETL
SEG .....XXXXXXXXXXXXXXXXXXXX
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG IQQLKEELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYAKIKELETILETQKVERSHS
SEG .....
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG AKLEQDILEKESIILKLERNLKEFQHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMK
SEG .....
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG HLLQLKEEEEETNRQETEKLEELSASSARTQNLKADLQKKEEDYADLKEKLTDAKKQIK
SEG .....XXXXXXXXXXXXXXXXXXXX
COILS CCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG QVQKEVSVMRDEDKLLRIKINELEKKRNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQY
SEG .....
COILS CCCCCCCCCC.....
3kar- .....

SEG ERACKDLNVKEKIIEDMRMTLEEQEQTQVEQDVLEAKLEEVERLATELEKWKEKCNLE
SEG .....XXXXXXXXXXXXXXXXXXXX
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar- .....

SEG TKNNQRSNKEHENNTDVLGKLTNLQDELQEQKYNADRKKWLEEKMLITQAKEAENIR
SEG .....
COILS CC.....
3kar- .....

SEG NKEMKKAEDRERFFKQONEMEILTAQLTEKSDLOKQWREERDQLVAALEIQLKALISSN
SEG .....

```

```

COILS .....
3kar- .....

SEQ      VQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGS
SEG      .....
COILS    .....
3kar-    .....

SEQ      VVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVKHGPCTTPVTVEIPKARKRKSNE
SEG      .....
COILS    .....
3kar-    .....

SEQ      MEEDLVKCNKKKATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKKTYSLRSQASIIGVN
SEG      .....
COILS    .....
3kar-    .....

SEQ      LATKKKEGTLQKFGDFLQHSFSLQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKL
SEG      .....
COILS    .....
3kar-    .....

SEQ      YTSEISSPIDISGQVILMDQMKESDHQIIKRRLRTKTAK
SEG      .....
COILS    .....
3kar-    .....

```

Prosites for DKFZphtes3_35b4.3

PS00017 152->160 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_35b4.3

```

HMM_NAME      Kinesin motor domain

HMM            *RCRPlNeReindgcscvVQWPpWtGyktvhngheds.....phks
               R+RP+ + E++ + +V + +++++ ++ +
Query         64' RIRPFTQSEKELESEGCVHILDSQTVVLLKEPQCILGRLSEKSSGQMAQK 112

HMM            FtFDHVFwWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQTGSGKTYTM
               F+F +VF++++TQ++ +++ + V+D+++G IF+YG T SGKTYT
Query         113 FSPSKVFGPATTKQEFFQGCIMQPVKDLLKGQSRILFTYGLTNSGKTYTF 162

HMM            MGpgggehPDHmGIIPRCCHDIFdrIdkfgekDhdFW.....
               G +++GI+PR+++ +FD++ + +++
Query         163 QG----TEENIGILPRTLNVLFDSLQERL-YTKMNLKPHRSREYLRLSSE 207

HMM            .....
Query         208 QEKEEIASKSALLRQIKEVTVHNDSDDTLYGLTNSLNISEFEESIKDYE 257

HMM            .....hvkCSYMEIYNEeiYDLLCPnP...qhMkpLnIHEHPN
               +V +S++EIYNE+IYDL +P++ Q++K L++ + +
Query         258 QANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKMLRLSQDVK 307

HMM            MGpYVqGCTEfHvCSYeDachWIWqGnknRHVAaTnMndhSSRShtIFTI
               ++++++ V +A +++ +G K+ VA T++N SSRShtIFT+
Query         308 GYSFIKDLQWQVSDSKEAYRLLKLGIKHQSVAFKLNASSRShtIFTV 357

HMM            HVeQrHk.qcdehvcHSKMNLVDLAGSERVnrTGAEGQRlKEGcNINqSL
               ++ Q + + +++S ++L DLAGSER+ +T+ EG RL+E +NIN SL
Query         358 KILQIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGERLRETGNINTSL 407

HMM            ttLGnVInaLaDgqTKYmYggghGHIPYRDSKLTWLLQDSLGGNcKtCMIA
               +TLG++IN+L + + + +H+P+R+SKLT+ +Q + G +K CMI+
Query         408 LTLGKCINVLKNS---KSKFQQHVFPRESKLTHTYFQSFFNGKGKICMIV 454

HMM            CIWPadWNYEETLSTLRYAdRAKnIkNkPQINEDPca*
               +I+ + Y+ETL++L++ + A+++ + ++N+++++
Query         455 NISQCYLAYDETLNVLFKFSIAIAQRVCVPDTLNSQDK 491

```


DKFZphtes3_35b5

group: metabolism

DKFZphtes3_35b5 encodes a novel 466 amino acid protein, with similarity to bovine accessory subunit for vacuolar ATPase and rat C7-1 protein.

The vacuolar proton-ATPase (V-ATPase) translocates protons into intracellular organelles or across the plasma membrane of specialized cells. The catalytic domain consists of a hexamer of 3 A subunits and 3 B subunits, plus accessory subunits C, D, and E. The rat homolog C7-1 seems to be enriched in aged adult rats in the frontal cortex.

The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A

complete cDNA, complete cds potential start at Bp 8, EST hits matches perfect to I54197 hypothetical protein, but possess 186 aa additional at N-terminus

Sequenced by DKFZ

Locus: unknown

Insert length: 2043 bp

Poly A stretch at pos. 2033, polyadenylation signal at pos. 2012

```
1  GCGCGCCATG GCGACGGCTC GAGTGGCGAT GGGGCCGCGG TGGCGCCAGG
51  CGCTCTGGCG CATGCCGTGG CTGCCGGTGT TTTTGTCTTT GCGCGCGGCG
101 GCGCGCGGCG CAGCGGCGGA GCAGCAGGTC CCGCTGGTGC TGTGGTCGAG
151 TGACCGGGAC TTGTGGGCTC CTGCGGCCGA CACTCATGAA GGCCACATCA
201 CCAGCGACTT GCAGCTCTCT ACCTACTTAG ATCCCGCCCT GGAGCTGGGT
251 CCCAGGAATG TGCTGCTGTT CCTGCAGGAC AAGCTGAGCA TTGAGGATTT
301 CACAGCATAT GCGCGTGTGT TTGGAACAA GCAGGACAGC GCCTTTCTTA
351 ACCTAGAGAA TGCCCTGGAC CTGGCCCCCT CCTCACTGGT GCTTCCTGCC
401 GTCGACTGGT ATGCAGTCAG CACTCTGACC ACTTACCTGC AGGAGAAGCT
451 CGGGGCCAGC CCCTTGCATG TGGACCTGGC CACCCTGCGG GAGCTGAAGC
501 TCAATGCCAG CCTCCCTGCT CTGCTGCTCA TTCGCCTGCC CTACACAGCC
551 AGCTCTGGTC TGATGGCACC CAGGGAAGTC CTCACAGGCA ACGATGAGGT
601 CATCGGGCAG GTCCTGAGCA CACTCAAGTC CGAAGATGTC CCATACACAG
651 CGGCCCTCAC AGCGGTCCGC CCTTCCAGGG TGGCCCGTGA TGTAGCCGTG
701 GTGGCCGGAG GGCTAGGTCG CCAGCTGCTA CAAAAACAGC CAGTATCACC
751 TGTGATCCAT CCTCCTGTGA GTTACAATGA CACCGCTCCC CGGATCCTGT
801 TCTGGGCCCA AAATTCTCT GTGGCTTACA AGGACCAGTG GGAGGACCTG
851 ACTCCCTTCA CCTTTGGGGT GCAGGAAGTC AACCTGACTG GCTCCTTCTG
901 GAATGACTCT TTTGCCAGGC TCTCACTGAC CTATGAACGA TCTTTTGTA
951 CCACAGTGAC ATTCAAGTTC ATTCTGGCCA ACCGCCTCTA CCCAGTGTCT
1001 GCGCGGCACT GGTTTACCAT GGAGCGCCTC GAAGTCCACA GCAATGGCTC
1051 CGTCGCCTAC TTCAATGCTT CCCAGGTCAC AGGGCCAGC ATCTACTCCT
1101 TCACTGCGA GTATGTCAGC AGCCTGAGCA AGAAGGGTAG TCTCCTCGTG
1151 GCGCGCACGC AGCCCTCTCC CTGGCAGATG ATGCTTCAGG ACTTCCAGAT
1201 CCAGGCTTTC AACGTAATGG GGGAGCAGTT CTCCTACGCC AGCGACTGTG
1251 CCAGCTTCTT CTCCCCCGGC ATCTGGATGG GGCTGCTCAC CTCCTGTTC
1301 ATGCTCTTCA TCTTCACCTA TGGCCTGCAC ATGATCCTCA GCCTCAAGAC
1351 CATGGATCGC TTTGATGACC ACAAGGGCCC CACTATTCTT TTGACCCAGA
1401 TTGTGTGACC CTGTGCCAGT GGGGGGGTTG AGGGTGGGAC GGTGTCCGTG
1451 TTGTTGCTTT CCCACCCTGC AGCGCACTGG ACTGAAGAGC TTCCCTCTTC
1501 CTACTGCAGC ATGAAGTCA AGCTCCCTC AGCCATCTT GCTCCCTCTT
1551 CAGCCCGCTG AGGAGCTTTC TTGGGCTGCC CCCATCTCTC CCAACAAGGT
1601 GTACATATTC TGGTAGATG CTAGACCAAC CAGCTTCCCA GGGTTCGTG
1651 CTGTGAGGCG TAAGGGACAT GAATTCTAGG GTCTCCTTTC TCCTATTATA
1701 TTCTTGTGGC TACATCATCC CTGGCTGTGG ATAGTGCTTT TGTGTAGCAA
1751 ATGCTCCCTC CTTAAGGTTA TAGGGCTCCC TGAGTTTGGG AGTGTTGAAG
1801 TACTACTTAA CTGTCTGTCC TGCTTGGCTG CCGTTATCGT TTTCTGGTGA
1851 TGTGTGTGTA ACAATAAGAA GTACACGGGT TTATTCTGTG GGCCTGAGAA
1901 GGAAGGGACC TCCACGACAG GTGGGCTGGG TGCGATCGCC GGCTGTTTGG
1951 CATGTTCCCA CCGGGAGTGC CGGGCAGGAG CATGGGGTGC TTGGTTGTTT
2001 CCTTCCTAAT AAAATAAAGC CGGGTCGCCA TGCAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

95014142:
A novel accessory subunit for vacuolar H(+)-ATPase from chromaffin granules.

97215246:
Identification of a rat brain gene associated with aging by PCR differential display method.

Peptide information for frame 2

ORF from 8 bp to 1405 bp; peptide length: 466
Category: strong similarity to known protein

```

1 MATARVRMGP RCAQALWRMP WLPVFLSLAA AAAAAAAEQQ VPLVLWSSDR
51 DLWAPAADTH EGHITSDQLQ STYLDPALEL GPRNVLLFLQ DKLSIEDFTA
101 YGGVFGNKQD SAFSNLENAL DLAPSSLVLP AVDWYAVSTL TTYLQEKLGA
151 SPLHVDLATL RELKLNASLP ALLLIRLPYT ASSGLMAPRE VLTGNDEVIG
201 QVLSTLKSED VPYTAALTAV RPSRVARDVA VVAGGLGRQL LQKQPVSPVI
251 HPPVSYNDTA PRILFWAQN FSVAYKDQWED LTPLTFGVQE LNLTSFVND
301 SFARLSLTYE RLFGTTFVK FILANRLYPV SARHWTMER LEVHSNGSVA
351 YFNASQVTGP SIYSFHCEYV SLSKKGSLV VARTQPSWPQ MMLQDFQIQA
401 FNMGEQFSY ASDCASFFSP GIWMGLLTSL FMLFIFTYGL HMILSLKTMD
451 RFDDHKGPIT SLTQIV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35b5, frame 2

TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds., N = 1, Score = 2088, P = 3.8e-216

PIR:A55116 vacuolar ATPase (EC 3.6.1.-) chain Ac45 - bovine, N = 1, Score = 2011, P = 5.5e-208

PIR:I54197 hypothetical protein - human, N = 1, Score = 1464, P = 5.1e-150

>TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds.
Length = 463

HSPs:

Score = 2088 (313.3 bits), Expect = 3.8e-216, P = 3.8e-216
Identities = 408/463 (88%), Positives = 426/463 (92%)

```

Query:      4 ARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAEQQVPLVLWSSDRDLWAPAADTHEGH 63
              +R+R G R A  LW      + LSL A AAA AAEQQVPLVLWSSDRDLWAP ADTHEGH
Sbjct:      8 SRIRTGTRWAPVLW-----LLLSLVAVAAVAAEQQVPLVLWSSDRDLWAPVADTHEGH 61

Query:     64 ITSDLQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 123
              ITSD+QLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA
Sbjct:     62 ITSDMQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 121

Query:    124 PSSLVLPVADWYAVSTLTITYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 183
              PSSLVLPVADWYA+STLTITYLQEKLGASPLHVDLATL+ELKLNASLPALLLIRLPYTASS
Sbjct:    122 PSSLVLPVADWYAISTLTITYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 181

Query:    184 GLMAPREVLTGNDEVIGQVLSTLKSEDVPYTAALTAVRPSRVARDVAVVAGGLGRQLLQK 243
              GLMAPREVLTGNDEVIGQVLSTL+SEDPYTAALTAVRPSRVARDVA+VAGGLGRQLLQ
Sbjct:    182 GLMAPREVLTGNDEVIGQVLSTLESEDPYTAALTAVRPSRVARDVAVVAGGLGRQLLQ 241

Query:    244 QPVSPVIHPPVSYNDTAPRILFWAQNFSVAYKDQWEDLTPLTFGVQELNLTGSFVNDSEFA 303
              Q SP IHPPVSYNDTAPRILFWAQNFSVAYKD+W+DLT LTFGV+ LNLTSFVNDSEFA
Sbjct:    242 QVASPAIHPPVSYNDTAPRILFWAQNFSVAYKDEWKDLTSLTFGVENLNLTSFVNDSEFA 301

Query:    304 RLSLTIERLFGTTVTFKIFILANRLYPVSARHWTMERLEVHSNGSVAYFNASQVTGPSIY 363
              LSLTYE LFG TVTFKIFILA+R YPVSAR+WFTMERLE+HSNGSVA+FN SQVTGPSIY

```

Sbjct: 302 MSLTYEPLFGATVTFKFIASREYPVSARYWFTMERLEIHSNGSVAHFNVSQVTGPSIY 361
 Query: 364 SFHCEYVSSLSKKGSLVARTQPSPWQMLQDFQIQAFNVMGQFSYASDCASFFSPGIW 423
 SFHCEYVSSLSKKGSLV PS WQM L +FQIQAFNV GEQFSYASDCA FFSPGIW
 Sbjct: 362 SFHCEYVSSLSKKGSLVTNV-PSLWQMTLHNFIQAFNVGTGEQFSYASDCAGFFSPGIW 420
 Query: 424 MGLLTSLFMLFIFTYGLHMLSLKTMDFDDHKGPTISLTQIV 466
 MGLLT+LFMLFIFTYGLHMLSLKTMDFDD KGPTI+LTQIV
 Sbjct: 421 MGLLTTLFMLFIFTYGLHMLSLKTMDFDDHKGPTITLTQIV 463

Pedant information for DKFZphtes3_35b5, frame 2

Report for DKFZphtes3_35b5.2

[LENGTH] 466
 [MW] 51621.44
 [PI] 5.73
 [HOMOL] TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1
 protein (C7-1) mRNA, complete cds. 0.0
 [PIRKW] hydrolase 0.0
 [PROSITE] MYRISTYL 7
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 7
 [KW] SIGNAL PEPTIDE 38
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 11.59 %

SEQ MATARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAEQQVPLVLWSSDRDLWAPAADTH
 SEGXXXXXXXXX.....
 PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhhhcccccccccccccccccc
 MEM

SEQ EGHITSDLQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENAL
 SEG
 PRD cccccchhhhhccccccccccccccccccccccccccccccccccccchhhhhhhcc
 MEM

SEQ DLAPSSVLPAVDWYAVSTLTYYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYT
 SEGXXXXXXXXXXXXXXXXX...
 PRD ccc
 MEM

SEQ ASSGLMAPREVLGTGNDVIGQVLTSLKSEDVPTAALTAVRPSRVARDVAVVAGGLGRQL
 SEGXXXXXXXXXXXXXXXXXXXXXXXXX...
 PRD ccc
 MEM

SEQ LQKQPVSPVIHPPVSYNDTAPRILFWAQNFVAYKDQWEDLTPLTFGVQELNLTGSFWND
 SEG
 PRD hhhhhcc
 MEM

SEQ SFARLSLTYERLFGTIVTFKFIANRLYPVSARHWFTMERLEVHNSGVSAYFNASQVTGP
 SEG
 PRD hhhhhhhhhhhcc
 MEM

SEQ SIYSFHCEYVSSLSKKGSLVARTQPSPWQMLQDFQIQAFNVMGQFSYASDCASFFSP
 SEGXXXXXXXXX.....
 PRD ccc
 MEMMMMMMM

SEQ GIWMGLTSLFMLFIFTYGLHMLSLKTMDFDDHKGPTISLTQIV
 SEG
 PRD cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc
 MEM MMM

Prosites for DKFZphtes3_35b5.2

PS00001	166->170	ASN_GLYCOSYLATION	PDOC00001
PS00001	257->261	ASN_GLYCOSYLATION	PDOC00001
PS00001	269->273	ASN_GLYCOSYLATION	PDOC00001

PS00001	292->296	ASN_GLYCOSYLATION	PDOC00001
PS00001	299->303	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00001	353->357	ASN_GLYCOSYLATION	PDOC00001
PS00004	375->379	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	159->162	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	331->334	PKC_PHOSPHO_SITE	PDOC00005
PS00005	374->377	PKC_PHOSPHO_SITE	PDOC00005
PS00005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	114->118	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00007	207->214	TYR_PHOSPHO_SITE	PDOC00007
PS00008	102->108	MYRISTYL	PDOC00008
PS00008	103->109	MYRISTYL	PDOC00008
PS00008	200->206	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35b5.2)

DKFZphtes3_35e21

group: differentiation/development

DKFZphtes3_35e21.2 encodes a novel 104 amino acid putative interleukin precursor, related to interleukin-7.

Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells.

This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

similarity to interleukin-7 precursor

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2095 bp

Poly A stretch at pos. 2085, polyadenylation signal at pos. 2067

```
1 GGATGAAAGT GATTTAATTC ATTTTATAGAA TTTT TTTT TTTT GTTTTGT TTT
51 AGCAACATGC TGAACAACCTA ATTTACTTTA AAAATAAGCC AGTTAAAACA
101 AAGGACCGTA AGCCCAAGTG GGGGGCAATA TTAGTCAGGA TCTTTGGGGT
151 CTAATTCAG ACCAACTTTC AGAAGCACTT CTTTGTCTCT GTTCTCACCT
201 CTGCTGTCCC TCTCTTCCCT CATCCCCTAA GAGAGACAAA GATAAAAGCC
251 CACCTGCATC CCTAAGTCTT ACTGAGATCA GCCACCCAG GGGAGAGAAA
301 CTGGATCTAC TTACAGCCAC CCCCTGTTTC CATCCATATA CTTACTTCCC
351 CCAATTTGCA TGTGATTATG GAAACAAGTC ATGCTCATGA AAGCAACTGT
401 AAAATAAAAG GTTATGGAGT AGTTCAGCAA CTTCTTCACA GCCAGCTTTG
451 TGGAGCTGGG GAGGACTTAG GGCCCATTTG AGTCTCTTAT GTGTACAGCT
501 TCAGGGCTGT CCCTTTCAGT TTGATTTTAA GCAATGCCTC ACTTCATAGC
551 TTAGGGGGTA AGGATTCCAT TCAGGTAGGT TGTCTAAAGG AACTAATGGG
601 ACCTCTCAGT GAATTAGCTG ACCAGATTTT AGGAAATCTT TTTAATTTCT
651 ATGATTTTCC TTCTCACATT TTGAAATGGT AAAATTGACT GGAAATAATT
701 TTTCTTGGTG CCTTATTGGT TTTCTTGGCA AACCTTTCTC ATATTTTCTC
751 ATGACCATTG CCACTGACCA AGGCCCATGT GTGTGTTGTG TGTAATTTGT
801 GGCATGTACA AGCTTAAATA ACGTGCCGAC AGCACTGTTT CAAAGTTGGT
851 ATTCAATTAG CTGTTGCCTC CTGGGCTGGA GCTGCGCTAA TCCTGACACC
901 GGCTGCCAGG AGAAAACCTC ATGGATCACA CACCAACCTT TAATAACAGC
951 ATCCGTGACC TGCACCTCTC AGTACAGAAAT GGGAAACCCA GAGCTAGGAA
1001 ATGTAGTTGT ATATTTTAAT GAATGCTAC CCCAGCCAAA GAAGCTTCTT
1051 TCATTTTGTG GCTCTACAGA AAGCCCAAGG GGGGTAGGAG GGACAGAGCT
1101 TTGAATAAAT GCTTTCTAAC ACTAAATGTG GCCAACAGGA CAGAGCACAT
1151 CACACGTATA GGCAGGTGTG AGGGACAGTG GCTAAGAATT GCCTGCTCCC
1201 TCTGCATGCT CTTTCTTGT TCCAAAGTCC AATCAAGTGA TCCTGGGAAA
1251 CAAATCTGTC TGGATTGCGG AGGGTGGTTC TGAAAGAACT GCCAAGACGT
1301 TAAAGAAGGG TGAAGAGTAG GCAGAAATATA AGTAGCTAAC CTGAGTCAAG
1351 ACTCTCAAAA GCTAGCAGCC TGATGACAAT AGGATTTATT TCAGCCAGGA
1401 TAGTGTCTGT CTGTGAGTGC ATCATTTTAA GACAGTATGA CTTCATGTTG
1451 TTACAAAATA TGTATAGTAT GTATGTTTTG TGGGTTGTAT ATATACATAA
1501 TATATATTAT ATATATATAT GAGAGATTG GTGACTTTT ATACGGGTTT
1551 GGTGCAGGTG AATTTATTAC TGAGCCAAAT GAGGCACATA CCGAGTCAGT
1601 AGTTGAAGTC CAGGGCATTC GATACTGTTT ATGATTTCCA TATATGTATA
1651 GTGCCTATCC CATGCTGTAG TCACTGTTAT GTTAAATCCA GAAGTTACAC
1701 TAGAGCCAGC GATACTTTAT TTGTAGACAA TCAATTGAA TCCATATGTT
1751 ATTACTGGCA GATGATACAT GATTACAGTT CTGAATCTGT AACACTTACA
1801 AAAGGAAACC CAGAGCAGCT TGATGAGTTT TTGTTTCTGC TTCGTTCCCTG
1851 GGAGTCAGTA GAAACAGCAG TTGTATGTGG TTATGTTAGT CTCAGATAC
1901 TTAATTTGTT GACCTTACTT CAGAAAAATT TTGTATGTAT TATATTTGTG
1951 GGAAGGTAAA ATAATCATTG GAGATTTTAA TCAAAATATGA AGATTAGTTA
2001 TTTATGAAAA ACAAAGAAAT GTCTATTTT CTTTGTTCCT AATTAATGTA
2051 GATAAATTTT AAAATGCATT AAAGTAATGG TCCGGAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

89098903:
Human interleukin 7: molecular cloning and growth factor
activity on human and murine B-lineage cells.

Peptide information for frame 2

ORF from 368 bp to 679 bp; peptide length: 104
Category: similarity to known protein

1 METSHAHESN CKIKGYGVVQ QLLHSQLCGA GEDLGPIGVS YVYSFRAVPF
51 SLILSNASLH SLGGKDSIQV GCLKELMGPL SELADQILGN LFNFYDFPSH
101 ILKW

BLASTP hits

Entry B32223 from database PIR:
interleukin-7 precursor (clone 1) - human
Score = 66, P = 7.0e-01, identities = 21/70, positives = 33/70

Alert BLASTP hits for DKFZphtes3_35e21, frame 2

PIR:B32223 interleukin-7 precursor (clone 1) - human, N = 1, Score =
66, P = 0.72

TREMBL:PADAL1_1 gene: "dall"; P.abies dall mRNA, N = 2, Score = 59, P
= 0.77

PIR:C32223 interleukin-7 precursor (clone 4) - human, N = 1, Score =
66, P = 0.79

TREMBL:PRU76726_1 gene: "PrMADS3"; product: "MADS-box protein"; Pinus
radiata MADS-box protein (PrMADS3) mRNA, complete cds., N = 2, Score =
59, P = 0.94

>PIR:B32223 interleukin-7 precursor (clone 1) - human
Length = 133

HSPs:

Score = 66 (9.9 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 21/68 (30%), Positives = 33/68 (48%)

Query: 39 VSYVYSFRAVPFSLIL-----SNASLHSLGGK--DSIQVGCLKELMGPLSELADQILGNL 91
VS+ Y F P L+L S+ + GK +S+ + + +L+ + E+ L N
Sbjct: 4 VSFRIYFGLPPLILVLLPVASSDCDIEGRDGRQYESVLMVSIQQLDSMKEIGSNCLNNE 63

Query: 92 FNFYDFPSHI 101
FNF F HI
Sbjct: 64 FNF--FKRHI 71

Pedant information for DKFZphtes3_35e21, frame 2

Report for DKFZphtes3_35e21.2

[LENGTH]	104
[MW]	11339.12
[pI]	5.87
[PROSITE]	MYRISTYL 2
[PROSITE]	PKC_PHOSPHO_SITE 1
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	Alpha_Beta

SEQ METSHAHESNCKIKGYGVVQQLLHSQLCGAGEDLGPIGVS YVYSFRAVPFSLILSNASLH
PRD ccchhhhhccccccccchhhhhhhhhhhccccccccceeeeeccccceeeeecccccc

SEQ SLGGKDSIQVGCLKELMGPLSELADQILGNLFNFYDFPSHILKW
PRD cccccceccccccccccccchhhhhhhcccccccccccccccc

Prosites for DKFZphtes3_35e21.2

PS00001	56->60	ASN_GLYCOSYLATION	PDOC00001
PS00005	44->47	PKC_PHOSPHO_SITE	PDOC00005
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35e21.2)

DKFZphtes3_35g6

group: testes derived

DKFZphtes3_35g6 encodes a novel 482 amino acid protein with high partial similarity to H. sapiens chromosome 19, cosmid R27216.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to R27216_1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="15"

Insert length: 3177 bp

Poly A stretch at pos. 3167, polyadenylation signal at pos. 3148

```
1 GGAGGCAGCG CCGGCCTCCG GAGGCGGCCT GGGCGATGGC GCGGGAGTTT
51 TGTCCATAAC CTGGGCAACC GCGCAGCTGG AGGATGGCCT CACTCGGGCC
101 TCCCGCAGCT GGGGAGCAGG CGTCGGGGGC TGAGGCGGAG CCGGGCCCCG
151 CGGGGCCGCC GCCCGCGCCC TCACCGTCTT CTCTGGGGCC CTTGCTCCCC
201 CTGCAGCGGG AACCTCTCTA CAACTGGCAG GCGACCAAGG CGTCGCTGAA
251 GGAGCGCTTC GCCTTCTCTT TCAACTCGGA GCTGCTGAGC GATGTGCGCT
301 TCGTACTGGG CAAGGGTCTG GCGCGCGCGG CCGCTGGGGG CCCGACGCGC
351 ATCCCCGCCC ACCGCTTCGT GCTGGCGGCC GGCAGCGCCG TCTTTGACGC
401 CATGTTCAAC GCGGCGATGG CCACCACGTC GGCCGAGATC GAGCTGCCGG
451 ACGTGGAGCC CGCAGCCTTC CTGGCGCTGC TGAGATTCTT ATATTGAGAT
501 GAAGTTCAAA TTGGTCCAGA AACAGTTATG ACCACTCTTT ATACTGCCAA
551 GAAATACGCA GTCCCAGCCT TGGAAACACA CTGTGTAGAA TTTCTCACCA
601 AACATCTTAG GGCAGATAAT GCCTTTATGT TACTTACTCA GGCTCGATTA
651 TTTGATGAAC CTCAGCTTGC TAGTCTTTGT CTAGATACAA TAGACAAAAG
701 CACAATGGAT GCAATAAGTG CAGAAAGGTT TACTGATATT GATATAGATA
751 CACTCTGTGC AGTTTATAG AGAGACACAC TCAGTATTCG AGAAAGTCGA
801 CTTTTTGGAG CTGTTGTACG CTGGGCAGAA GCAGAATGTC AGAGACAACA
851 ATTACCTGTG ACTTTTGGGA ATAAACAAAA AGTTCTAGGA AAAGCATTCT
901 CCTTAATCCG GTTCCCACTG ATGACAATTG AGGAATTTCG AGCAGGTCCT
951 GCTCAATCTG GAATTTGTGC AGATCGTGAA GTGGTAAACC TCTTCTTCA
1001 TTTTACTGTC AACCCCTAAAC CCCGAGTTGA ATACATTGAC CGACCAAGAT
1051 GCTGTCTCAG GGGAAAGGAA TGCTGCATCA ATAGATTCCA GCAAGTAGAA
1101 AGCCGCTGGG GTTACAGTGG GACGAGTGAT CGAATCAGAT TCACAGTTAA
1151 TAGAAGGATC TCTATAGTTG GATTTGGCTT GTATGGATCT ATTCATGGCC
1201 CTACAGATTA TCAAGTGAAT ATACAGATCA TTGAATATGA GAAAAAGCAA
1251 ACCCTGGGAC AGAATGATAC CGGCTTTAGT TGTGATGGGA CAGCTAACAC
1301 ATTCAGGGTC ATGTTCAAGG AACCCATAGA GATCCTGCCC AATGTGTGCT
1351 ACACAGCATG TGCAACACTC AAAGGTCCAG ATTCCCCTA TGGCACAAAA
1401 GGATTGAAGA AAGTAGTGCA TGAGACACCT GCTGCAAGCA AGACTGTTTT
1451 TTTCTTTTTT AGTTCCCTG GCAATAATAA TGGCACTTCA ATAGAAGATG
1501 GACAAATTC AGAAATCATA TTTTATACAT AATTTAGCAT TATAATACAT
1551 CTGGGCTAAA TAATACCATA CAATCTAGTG TCAAAAACAT AAATGGCCAC
1601 AAAAAAGTAG TTTGAGTGTT ATGAATATTT AAAATTGTAA GATAAGAAAC
1651 AGTTTCTTAG AGCAGATAGA AAAATGCTTA TTTAAATCTT TGCATGATTT
1701 AAAAAACAGAT TTTCCATTTT CTTACAACCT TAAGAGAAAA GAACTGGGTT
1751 TAATGGTTTA AAAAAAAGCA CAGCTTTTTC ACCTTCATCT TGTATAATT
1801 CATAGATTGG CTGACTTAGG GTCTTTCAAT AGTTTGGGAA TTGAAAGATT
1851 CTGTGTATAT ATAGCTAGTT TGGGTTTGT TTTGTTTTAA CTATTTTGAA
1901 GGTTAGGTGA GATGGGCAAA TAGGCTTAAC TATTTTGAAG GTTGGATGAA
1951 AAGAGATGGG TCAGTATTC TACAGAATTC TTATTAACCT AAATAACTAA
2001 ATTTACAGAA ATTAAGAAGC TGACTTTATA TTTGGTGGTT TGAAGTATCT
2051 TGTGTGTAGC ATTTGTAATA ATGCTAAAAA AGGCCTAATA AAATGCCCAA
2101 GAAAAATATC AGTGCAATTA TAGAGAAGGA TATTTGTAG TAGTATAGTA
2151 ATGTGTTATG TAGTACAGTT TTAAGCTAT AAATGGAATT TTGTGTAAT
2201 TCACAAAAAT GTGATATAAA CAGGATCTAA GACTGGATTC CCTGTCACCTA
2251 AACTGCACCA CTATACCTGT CTCTCTGTGT GGGGGACACT GCTGATGATT
2301 CCCAAGATTG AGATGATGAC GGTGATGACG ACTGGGTGAA CAGGCATCAC
2351 TTCAACATTG TGATAATCCT TCACAGCAAG AAACCGAATA AAATACTAAC
2401 ATTTCTAACA ACTGCTCTGA CATTGTAAA AGATCCAACA GAATCACTCC
2451 TGTCTGAAAA TACGCTTTCT GCCACCTACA CATTTCTATT TAGGAAGTAA
2501 AATTTGCTTC ATGGTCATGA CCCCATTAGT CAGTGTTACA GCTGTGTTGG
2551 GGATAGGAAG TATATCTGGC AGATTGACAT TTATACACTT TTTTATAAAG
2601 CAGATTTTAA AATATAGTAA CATCCATTTT TTTCCCTTGA AAGTGATTCT
2651 CTTATAAAAA ATGAAAGTGG AGTTTAAGGT ATATCAAATC GTTGTGGAAG
2701 GTGATTAATA ATCAAAATTC TTTTAAATAT CAACTTAATT TTTTCTAAGT
```



```

2751 AAGATACAAA AAATTTTCAT CTAAAGTAAT ATTTCACTTT ATATTGTAAA
2801 GAAGGTAGGT ATATTGGTGG CTGAGGTCTC TTGAAATTGC TAAAGGGAAA
2851 TTTTCTATG GTAATGCTCT TACGGATATA AGCCTCAGTT AAATGGAATT
2901 ATCTATGGGA TGTGTGGTTC TGGTTAACTA AAAATTAACC AGTAAACACT
2951 CTGTAGTAAC CATTACAGAA AATACTTCTG CCTTAAAAAA TATGATATGC
3001 CAGAGATGAG TTAGTGTTTC TTGACGTTGG AGACCTATAA ATGCCTCATC
3051 TGTGTACTG AACAATTGAA ACTGCATGCA GCCATAAAG GGACAAGAAA
3101 CAGAACTGTT TACTAACTTT GGGACATCCC CTGGAGTTTT TAAAAATAAA
3151 TAAATATATA TATATATAAA AAAAAA

```

BLAST Results

Entry G37753 from database EMBL:
 SHGC-63477 Human Homo sapiens STS genomic.
 Score = 1627, P = 3.0e-66, identities = 327/329

Entry G37752 from database EMBL:
 SHGC-63476 Human Homo sapiens STS genomic.
 Score = 1578, P = 6.2e-64, identities = 320/324

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 84 bp to 1529 bp; peptide length: 482
 Category: similarity to unknown protein

```

1 MASLGPAAG EQASGAEEAP GPAGPPPPPS PSSLGPLLPL QREPLYNWQA
51 TKASLKERFA FLNSELLSD VRFVLGKGRG AAAAGGPORI PAHRFVLAAG
101 SAVFDAMFNG GMATTSAEIE LPDVEPA AFL ALLRFLYSDE VQIGPETVMT
151 TLYTAKKYAV PALEAHCVEF LTKHLRADNA FMLLTQARLF DEPQLASLCL
201 DTIDKSTMDA ISAEGETDID IDTLCAVLER DTLISRESRL FGAVVRWAEA
251 ECQRQQLPVT FGNKQKVLGK ALSLIRFPLM TIEEFAAGPA QSGILSDREV
301 VNLFLHFTVN PKPRVEYIDR PRCLRGKEC CINRFQOVES RWGYSGTSDR
351 IRFTVNRRIS IVGFGLYGS I HGPTDYQVNI QIIEYEKKQT LGQNDTGFSC
401 DGTANTFRVM FKEPIEILPN VCYTACATLK GPDSHYGTRG LKKVVHETPA
451 ASKTVFFFFS SPGNNGTSSI EDGQIPEIIF YT

```

BLASTP hits

Entry AC005306_2 from database TREMBL:
 product: "R27216_1"; Homo sapiens chromosome 19, cosmid R27216,
 complete sequence.
 Score = 1298, P = 1.9e-132, identities = 245/297, positives = 268/297

Entry CEF38H4_9 from database TREMBLNEW:
 gene: "F38H4.7"; Caenorhabditis elegans cosmid F38H4
 Score = 1237, P = 5.6e-126, identities = 248/446, positives = 322/446

Entry AC004678_1 from database TREMBL:
 product: "R34094_1"; Homo sapiens chromosome 19, cosmid R34094,
 complete sequence.
 Score = 555, P = 1.0e-53, identities = 112/137, positives = 123/137

Alert BLASTP hits for DKFZphtes3_35g6, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3_35g6, frame 3

Report for DKFZphtes3_35g6.3

```

[LENGTH] 482
[MW]      52771.47
[pI]      5.79

```

```

[HMOL]          TREMBL:AC005306.2 product: "R27216_1"; Homo sapiens chromosome 19, cosmid
R27216, complete sequence. 1e-142
[BLOCKS]        BL01075D Acetate and butyrate kinases family proteins
[SUPFAM]        POZ domain homology 3e-08
[SUPFAM]        A55R protein middle region homology 5e-06
[SUPFAM]        A55R protein 5e-06
[SUPFAM]        A55R protein carboxyl-terminal homology 5e-06
[PROSITE]       MYRISTYL 6
[PROSITE]       CAMP_PHOSPHO_SITE 2
[PROSITE]       CK2_PHOSPHO_SITE 9
[PROSITE]       TYR_PHOSPHO_SITE 1
[PROSITE]       PKC_PHOSPHO_SITE 7
[PROSITE]       ASN_GLYCOSYLATION 2
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY 11.20 %

```

WO 01/12659

PCT/IB00/01496

PS00008	392->398	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	463->469	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phtes3_35g6.3)

DKFZphtes3_35k16

group: metabolism

DKFZphtes3_35k16 encodes a novel 666 amino acid protein with weak similarity to fatty acid-CoA synthetases/ligases.

The novel protein contains a putative AMP-binding domain signature, which is present in enzymes, which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. Therefore it is a new fatty acid-CoA synthetase/ligase with unknown substrate.

The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

similarity to acyl-CoA synthetase

complete cDNA, complete cds, potential start codon at Bp 50, few EST hits, seems to be a testis specific cDNA, 5 of 6 EST hits are from testis derived libraries

Sequenced by DKFZ

Locus: unknown

Insert length: 2520 bp

Poly A stretch at pos. 2510, polyadenylation signal at pos. 2490

```
1 CAGATGTCCC AGCTCCAGTG CTGTGGAGCA TGGTTTCTGC ACACCTGGAA
51 TGACTGGAAC CCCAAAGACT CAAGAAGGAG CTAAAGATCT TGAAGTAGAC
101 ATGAATAAAA CAGAAGTTAC TCCCAGGCTG TGGACCACCT GTCGAGATGG
151 AGAAGTCCTT CTGAGGCTAT CCAAACACGG ACCAGGCCAT GAGACCCCGA
201 TGACCATCCC TGAATTTTTT CGAGAGTCAG TCAACCGATT TGGAACTTAT
251 CAGGCCCTCG CATCCAAGAA TGGCAAAAAG TGGGAAATTC TGAATTTCAA
301 CCACTACTAT GAGGCTGTGC GGAAGGCTGC AAAATCCTTG ATCAAGCTGG
351 GTTTGGAGCG TTTCCACGGA GTTGGTATCC TGGGGTTTAA CTCTGCAGAG
401 TGGTTTATCA CTGCTGTTGG TGCCATCCTA GCCGGGGGTC TTTGTGTTGG
451 TATTTATGCC ACCAACTCTG CCGAGGCTTG TCAATATGTC ATCACTCATG
501 CCAAAGTGAA CATCTTGCTG GTTGAGAATG ATCAACAGTT ACAGAAAATC
551 CTTTCGATTTC CACAGAGCAG CCTAGAGCCC CTAAAAGCGA TCATCCAGTA
601 CAGACTGCCA ATGAAGAAGA ACAACAACCT GTACTCTTGG GATGATTTC
651 TGGAACTTGG CAGAAGTATC CCTGACACCC AACTGGAGCA GGTTCATCGAG
701 AGCCAGAAGG CGAATCAATG CGCAGTGCTC ATCTACACTT CAGGGACCAC
751 AGGCATACCC AAGGGAGTGA TGCTCAGTCA TGACAACATC ACGTGGATTG
801 CAGGAGCAGT GACAAAGGAC TTTAAACTGA CAGACAAGCA TGAGACGGTG
851 GTTAGCTACC TCCCCTCAG CCATATTGCA GCACAGATGA TGGACATCTG
901 GGTACCCATA AAGATTGGGG CGCTCACATA CTTTGCTCAA GCAGATGCTC
951 TCAAGGGCAC CTTGGTAAAGT ACTCTAAAGG AGGTAAAACC TACTGTCTTC
1001 ATTGGAGTGC CTCAAATTTG GGAGAAGATA CATGAGATGG TGAAGAAAAA
1051 TAGTGCCAAG TCCATGGGCT TGAAGAAGAA GGCATTCTGT TGGGCAAGAA
1101 ACATTGGCTT CAAGGTCAAC TCAAAAAGAA TGTGGGGGAA ATATAATACT
1151 CCCGTGAGCT ACCGCATGGC TAAGACTCTC GTGTTCAGCA AAGTCAAGAC
1201 ATCCCTTGGC TTGGATCACT GTCACCTTTT TATCAGTGGG ACTGCGCCCC
1251 TCAACCAAGA GACTGCCGAG TTCTTTCTAA GCTTGGACAT ACCTATAGGC
1301 GAGTTGTATG GGTGAGTGA GAGCTCGGGA CCCACACGTA TATCCAACCA
1351 GAATAACTAC AGGCTTCTAA GCTGTGGCAA GATCTTGACT GGGTGTAAAG
1401 ATATGCTGTT CCAGCAGAAC AAGGATGGCA TTGGGGAGAT CTGCCTCTGG
1451 GGTAGGCACA TCTTCATGGG CTATCTGGAA AGTGAGACTG AAACATACAGA
1501 GGCCATCGAT GATGAAGGCT GGCTACACTC TGGGGATCTG GGCCAGCTGG
1551 ACGGTCTGGG TTTCTCTATG GTCACCGGCC ACATCAAAGA AATCCTTATC
1601 ACTGCTGGTG GTGAAATGT GCCCCCCATT CCTGTTGAGA CCTTGGTTAA
1651 GAAGAAGATC CCCATCATCA GTAACGCCAT GTTAGTAGGA GATAAACTGA
1701 AGTTTCTGAG CATGTTGCTG ACGCTGAAGT GTGAGATGAA TCAGATGAGC
1751 GGAGAACCCT TGGACAAGCT GAACTTCGAG GCCATCAACT TCTGTCGGGG
1801 TCTGGGCAGC CAGGCATCCA CCGTGACTGA GATGGTGAAG CAGCAAGACC
1851 CCCTGGTCTA CAAGGCCATC CAGCAAGGCA TCAATGCTGT GAACCAGGAA
1901 GCCATGAACA ATGCACAGAG GATTGAAAAG TGGGTCATCT TGGAGAAGGA
1951 CTTTTCATC TATGGTGGAG AGCTAGGTCC AATGATGAAA CTTAAGAGAC
2001 ATTTTGTAGC CCAGAAATAC AAAAAACAAA TTGATCATAT GTACCCTGTA
2051 CTGCTTTGAT GGAGCTGCTC TCAGCTGTTT TGATGCCTTC AGCAGGAAGA
2101 CCTCATTTGA ATAAGTGAAA TGCTGCTCTA GGTAAGAGCT CTCCTGCTG
2151 TTTTAAAGAA GCCACATTC TCATTGGTCA GTTCTTTGAT GTTTCGCTG
2201 TTGGAGAGGT GCTCCCTAGA AGAACCTGCC ATACGTTTCA AAGCAATAAA
2251 ATCACTGTAT ATCTTTCTAA GGACCTTCAA GTCATGACTC CAGGGAAGCC
2301 TATTGGGAAG TCTACTAAAA ACTGCCTGAT TTACAAGAAA GACCTGAAC
```

2351 TGTGGGCTCC CATTGATTT TTTTCTCCTC AGGGGACTCA GACATTAGAA
 2401 AGAAAAAGCC TCACAGATTT GAAGAACTGG ACCCCCAAAT CAACTCACCT
 2451 GCCTGGAAGC AACTGGGAAA CCCTTCCAAT AAGTCCTGAT AATAAAGCAC
 2501 TTCAGGTGCC AAAAAAAAAA

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2047 bp; peptide length: 666
 Category: similarity to known protein

1 MTGTPKTEQEG AKDLEVDMNK TEVTPRLWTT CRDGEVLLRL SKHGPGETP
 51 MTEPEFFRES VNRFGTYPAL ASKNGKKWEI LNFNQYYEAC RKAASLIKL
 101 GLERFHVGVI LGFNSAEWFI TAVGAILAGG LCVGIYATNS AEACQYVITH
 151 AKVNILLVEN DQQLQKILSI PQSSLEPLKA IIQYRLPMKK NNNLYSWDDF
 201 MELGRSIPDT QLEQVIESQK ANQCAVLIYT SGTGIPKGV MLSHDNITWI
 251 AGAVTKDFKL TDKHETVVSY LPLSHIAAQM MDIWWPIKIG ALTYFAQADA
 301 LKGTIVSTLK EVKPTVFIGV PQIWEKIHEN VKKNSAKSMG LKKKAFVWAR
 351 NIGFKVNSKK MLGKYNTFVS YRMATLVFS KVKTSLGLDH CHSFISGTAP
 401 LNQETAEEFL SLDIPIGELY GLSESSGPHT ISNQNNYRL SCGKILTCK
 451 NMLFQQNKDG IGEICLWGRH IFMGYLESET ETTEAIDDEG WLHSGDLGQL
 501 DGLGFLYVTG HIKEILITAG GENVPPIPVE TLVKKKIPII SNAMLVGDKL
 551 KFLSMLLTILK CEMNQMSGEP LDKLNFEAIN FCRGLGSQAS TVTEMVKQD
 601 PLVYKAIQGG INAVNQEAMN NAQRIEKWVI LEKDFSIYGG ELGPMMLKKR
 651 HFVAQKYKKQ IDHMYH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35k16, frame 2

TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens mRNA for KIAA0631 protein, partial cds., N = 1, Score = 1641, P = 8.9e-169

PIR:E70937 probable fadD15 - Mycobacterium tuberculosis (strain H37RV), N = 2, Score = 532, P = 3.6e-62

PIR:H64041 long-chain-fatty-acid--CoA ligase homolog - Haemophilus influenzae (strain Rd KW20), N = 2, Score = 486, P = 6.5e-59

>TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens mRNA for KIAA0631 protein, partial cds.
 Length = 634

HSPs:

Score = 1641 (246.2 bits), Expect = 8.9e-169, P = 8.9e-169
 Identities = 319/628 (50%), Positives = 440/628 (70%)

Query: 38 LRLSKHGPGETPMTIPEFFRESVNRFGTYPALASKNGKKWEILNFNQYYEACRKAASL 97
 LR+ P + P T+ F E++++G AL K KWE ++++QYY R+AAK
 Sbjct: 2 LRIDPSCP--QLPYTVHRMFYEALDKYGDIALGFKRQDKWEHISYSQYLLARRAAKGF 59

Query: 98 IKLGLERFHVGILGFNSAEWFITAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILL 157
 +KLGL++ H V ILGFNS EWF +AVG + AGG+ GIY T+S EACQY+ N+++
 Sbjct: 60 LKGLKQAHSAVAILGFNSPEWFFSAVGTVFAGGIVTGIYTTSSPEACQYIAYDCCANVIM 119

Query: 158 VENDQQLQKILSIPQSSLEPLKAIQYRLPM-KNNNNLYSWDDFMELGRSIPDTQLEQVI 216
 V+ +QL+KIL I L LKA++ Y+ P K N+Y+ ++FMELG +P+ L+ +I
 Sbjct: 120 VDTQKQLEKILKI-WKQLPHLKAVVIYKEPPPNKMANVYTMEEFMELGNEVPPEALDAII 178

Query: 217 ESQRANQCAVLIYTS GTTGIPKGVMLSHDNITWIA--GAVTKDFKLT-D-KHETVVSYLPL 273

```

++Q+ NQC VL+YTS GTTG PKGVMLS DNITW A G+ D + + + E VVSYLPL
Sbjct: 179 DTQQNQCCVLVYTS GTTGNGPKGVMLSQDNITWTARYGSQAGDIRPAEVQEQEVVVSYLPL 238
Query: 274 SHIAAQMDIWIWPIKIGALTYFAQADALKGTLVSTLKEVKPTVFIVGVPQIWEKIHVMVK 333
SHIAAQ+ D+W I+ GA FA+ DALKG+LV+TL+EV+PT +GVP++WEKI E +++
Sbjct: 239 SHIAAQIYDLWTGIQWGAQVCAEPDALKGSLVNTLREVEPTSHMGVPRVWEKIMERIQE 298
Query: 334 NSAKSMGLKKKAFVWVARNIGFKVNSKMLGKYNTPVSYRMAKTLVFSKVKTSGLDHC 393
+A+S +++K +WA ++ + N G P + R+A LV +KV+ +LG C
Sbjct: 299 VAAQSGFIRRKMLLWAMSVTLEQNLT-CPGSDLKPFTTTLADYLVLAQVQALGFACQK 357
Query: 394 FISGTAPLNQETAEEFFLSLDIPIGELYGLSESSGPHTISNQNNYRLSCGKILTGC 453
G AP+ ET FFL L+I + YGLSE+SGPH +S+ NYRL S GK++ GC+ L
Sbjct: 358 NFYGAAPMAETQHFFLGLNIRLYAGYGLSETSGPHFMSPPYNYRLYSSGKLVPGCRVKL 417
Query: 454 FQONKDGIGEICLWGRHIFMGYLESETETTEAIDDEGWLHSGDLGQDLGLFLYVTGHIK 513
Q+ +GIGEICLWGR IFMGYL E +T EAID+EGWLH+GD G+LD GFLY+TG +K
Sbjct: 418 VQDQAEIGIGEICLWGRITIFMGYLNMEDKTCEAIDEEGWLHTGDAGRLDADGFLYITGRK 477
Query: 514 EILITAGGENVPPPIPVETLVKKKIPISNAMLVGDKLFLSMLLTLCCEMNQMSGEPLDK 573
E++ITAGGENVPP+P+E VK ++PIISNAML+GD+ KFLSMLLTLC ++ + + D
Sbjct: 478 ELIITAGGENVPPVPIEEAVKMEPLIISNAMLIGDQRKFLSMLLTLCCTLDPTSDQTON 537
Query: 574 LNFEAINFRCRGLGSQASTVTVMVKQDPLVYKAIQGGINAVNQEAMNNAQRIEKWVILEK 633
L +A+ FC+ +GS+A+TV+E++++D VY+AI++GI VN A I+KW ILE+
Sbjct: 538 LTEQAVEFCQRVGSRRATTVSEIIEKKDEAVYQAIIEGIRRVNMNAAARPYHIQKWAILER 597
Query: 634 DFSIYGGEGLGPMMLKRRHFVAQKYKKQIDHMY 665
DFSI GGELGP MKLKR V +KYK ID Y
Sbjct: 598 DFSISGGEGLGPTMKLRLTVLEKYKGIIDSFY 629

```

Pedant information for DKFZphtes3_35k16, frame 2

Report for DKFZphtes3_35k16.2

```

[LENGTH] 666
[MW] 74344.97
[pI] 8.67
[HOMOL] TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens
mRNA for KIAA0631 protein, partial cds. 1e-176
[FUNCAT] i lipid metabolism [H. influenzae, HI0002] 2e-55
[FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.13 lipid and fatty-acid transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YER015w]
2e-29
[FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YMR246w]
2e-23
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YMR246w] 2e-23
[BLOCKS] BL00455
[SCOP] dlci_5.19.1.1.1 Luciferase [Firefly (Photinus pyralis)] 1e-49
[EC] 1.13.12.7 Photinus-luciferin 4-monooxygenase (ATP-hydrolysing) 9e-17
[EC] 6.2.1.3 Long-chain-fatty-acid--CoA ligase 4e-34
[EC] 5.1.1.11 Phenylalanine racemase (ATP-hydrolysing) 6e-08
[EC] 6.2.1.12 4-Coumarate--CoA ligase 8e-18
[PIRKW] duplication 6e-07
[PIRKW] phosphopantetheine 3e-12
[PIRKW] multifunctional enzyme 3e-06
[PIRKW] ligase 6e-08
[PIRKW] acid-thiol ligase 4e-34
[PIRKW] transmembrane protein 5e-22
[PIRKW] monooxygenase 9e-17
[PIRKW] hydrolase 4e-34
[PIRKW] peroxisome 9e-15
[PIRKW] antibiotic biosynthesis 3e-12
[PIRKW] isomerase 6e-08
[PIRKW] flavonoid biosynthesis 1e-17
[PIRKW] magnesium 9e-15
[PIRKW] ATP 5e-22
[PIRKW] oxidoreductase 9e-17
[PIRKW] liver 2e-31
[SUPFAM] alpha-aminoacyl-cysteine-valine synthetase 3e-07
[SUPFAM] human long-chain-fatty-acid--CoA ligase 4e-34
[SUPFAM] gramicidin S synthetase I 6e-08
[SUPFAM] peptide synthetase ppsE 7e-06
[SUPFAM] gramicidin S synthetase I repeat homology 3e-12
[SUPFAM] peptide synthetase ppsD 2e-07

```

```

[SUPFAM]    probable acyl-CoA ligase medium chain 2e-09
[SUPFAM]    acetate--CoA ligase 8e-10
[SUPFAM]    acetate--CoA ligase homology 4e-54
[SUPFAM]    surfactin synthetase 3e-12
[SUPFAM]    4-coumarate--CoA ligase 8e-18
[SUPFAM]    short-chain alcohol dehydrogenase homology 8e-07
[SUPFAM]    acyl carrier protein homology 2e-29
[PROSITE]   MYRISTYL      12
[PROSITE]   AMP_BINDING   1
[PROSITE]   AMIDATION     1
[PROSITE]   CAMP_PHOSPHO_SITE 1
[PROSITE]   CK2_PHOSPHO_SITE 9
[PROSITE]   TYR_PHOSPHO_SITE 3
[PROSITE]   PKC_PHOSPHO_SITE 10
[PROSITE]   ASN_GLYCOSYLATION 2
[PFAM]      AMP-binding enzymes
[KW]        Irregular
[KW]        3D
[KW]        LOW_COMPLEXITY 1.80 %

```

```

SEQ      MTGTPKQTQEGAKDLEVD MNKTEVTPRLWTTCDGEVLLRLSKHGPGHETPMTIPEFFRES
SEG      .....
llci-    .....

SEQ      VNRFGTYPALASKNGKKWEILNFNQYYEACRKAASLIKGLERFHGVGILGFNSAEWFI
SEG      .....
llci-    .....

SEQ      TAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQKILSIPQSSLEPLKA
SEG      .....
llci-    .....

SEQ      IIQYRLPMKKNNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIYTSGTTGIPKGV
SEG      .....
llci-    .....

SEQ      MLSHDNITWIAGAVTKDFKLTDKHETVVSYLPLSHIAAQMMDIWVPIKIGALTYFAQADA
SEG      .....
llci-    .....

SEQ      LKGTIVSTLKEVKPTVFIGVPQIWEKIHEMVKKNSAKSMGLKKKAFVWARNIGFKVNSKK
SEG      .....
llci-    .....

SEQ      MLGKYNTPVSYRMAKTLVFSKVKTSGLDCHCSFISGTAPLNQETAFFFLSLDIPIGELY
SEG      .....
llci-    .....TTTCEEETTTTCCCHHHHHHHHHCCCCBCEE

SEQ      GLSESSGPHITISNQNNYRLSCGKILTGCNNMLFQONKDGIGEICLWGRHIFMGYLESET
SEG      .....
llci-    ECGGGTTEEECCCCCEEEETTTTTEEEETTTTCEETTEEEETTTTCEETTTTHH

SEQ      ETTEAIDDEGWLHSGDLGQLDGLFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPII
SEG      .....XXXXXXXXXXXXX.....
llci-    HHHHHBTTTCEEEEEEEETTTTCEEE-----ECEEETTEEECHHHHHHHHHHT-TTE

SEQ      SNAMLVGDKLKFSLMLTLKCEMNQMSGEPLDKLNFEAINFCRGLGSQASTVTEMVKQOD
SEG      .....
llci-    EEEEEEE.....

SEQ      PLVYKAIQGINAVNQEAMNNAQRIEKWVILEKDFSIYGGELGPMMLKRHFVAQKYKKO
SEG      .....
llci-    .....

SEQ      IDHMYH
SEG      .....
llci-    .....

```

Prosites for DKFZphtes3_35k16.2

PS00001	19->23	ASN_GLYCOSYLATION	PDOC00001
PS00001	246->250	ASN_GLYCOSYLATION	PDOC00001
PS00004	332->336	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005

PS00005	308->311	PKC_PHOSPHO_SITE	PDOC00005
PS00005	335->338	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	558->561	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	173->177	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	478->482	CK2_PHOSPHO_SITE	PDOC00006
PS00006	591->595	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	658->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	597->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	3->9	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	303->309	MYRISTYL	PDOC00008
PS00008	387->393	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	586->592	MYRISTYL	PDOC00008
PS00009	74->78	AMIDATION	PDOC00009
PS00455	227->239	AMP_BINDING	PDOC00427

Pfam for DKFZphtes3_35k16.2

HMM_NAME	AMP-binding enzymes	
HMM	*TYRELNERANRLARHLRsekGlrPGDiVgIMMDRSMWMIVaMLGIWKAG + + +E +A L+ +G VGI+ +S + ++ G + AG	
Query	82 NFNQYYEACRRAAKSLI-KLGLERFHGVGILGFNSAEWFI TAVGAILAG	129
HMM	GAYVPIDPeYPdERiQYMLEDSGARLLITQrh....HmqRIPdemwvwdH G +V I +E QY++ ++ +L+++ + + IP++++ +	
Query	130 GLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQKILSIPQSSLEPLK	179
HMM	IiviDwe.....WddlWWHedeeNpqPwvdPeDLAYIIY +I++ + + +++++ + E ++ +++++ A +IY	
Query	180 AIIQYRLPMKKNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIY	229
HMM	TSGTTGKPKGVMIHrNiVnycqWmNWRyGmtedDRILWfTSDpYWFda TSGTTG PKGVM++H NI+ + + + +T+ + + + + + A	
Query	230 TSGTTGIPKGVMLSHDNITWAGAVTKDFKLTDKHETVVSYP-LSHIAA	278
HMM	SVWDMFWpLLnGaTLyIpPeEtRrDPerWWqYIqRHgITWwylTPSMFRM +++D++ P+ GA Y + ++ + +++++ +T+ ++P +++	
Query	279 QMMDIWVPIKIGALTYFAQADAL--KGTLVSTLKEVKPTVFIGVPQIWEK	326
HMM	LMpd.....	
Query	327 IHEMVKKNSAKSMGLKKAFVWARNIGFKVNSKKMLGKYNTPVSYRMAKT	376
HMMpsLRhVMFgGepLsPehWdWWRkrfgfkgRIINMYWPT ++ + +++G PL++E+++ ++ + ++I Y+ +	
Query	377 LVFSKVKTSGLDHCFSFISGTAPLNQETAFFFL-SLD--IPIGELYGLS	423
HMM	ETTVWtTwMrIiPdepeqWrwiPIGRPIpNTqWYIMDdnMQIQPiGViGE E++ T+ + + R +++G+ + + + +N G IGE	
Query	424 ESSGPHTISNQNN--Y---RLSCGKIILTCKNMLFQQN----RDG-IGE	463
HMM	LYIGWPGVARGYWNRPTELTEERfipNPFWPGEYRrGWNrMYRTGDLAR +++ G ++ GY+ + +T E+ + ++ ++GDL++	
Query	464 ICLWG-RHIFMGYLESETETTEAIDDEGW-----LHSGDLGQ	499
HMM	WlPDGnIEYLGRID.DQVKIRGYRIELGEIEhqlr.qHPgIqEAVV* + G+++ G I + G+++ + +E+ + ++P I+ A	
Query	500 LDGLGFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPIISNAML	545

DKFZphtes3_35k24

group: transmembrane protein

DKFZphtes3_35k24 encodes a novel 514 amino acid protein without similarity to known proteins.

The novel protein contains 5 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown ;

membrane regions: 5

Summary DKFZphtes3_35k24 encodes a novel 514 amino acid protein.

No homologues found in bacteria yeast and C.elegans, specific for mammals?

unknown

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2706 bp

Poly A stretch at pos. 2696, polyadenylation signal at pos. 2675

```

1 CCGTGTGTCAG TCGCCCCGCG CCCCAGCGCA CCCTTCGGGT AAACACGAA
51 CTGGGAGTTC TGAAGAATGG GTAAAGACTT TCCTTACTAT TTCCAGCATC
101 CCTGGTCTCG CATGATTGTG GCTTACTTGG TGATCTTCTT TAACCTCTTA
151 ATATTGTCGG AGGACCCAGT TTCTCATAGC CAAACAGAAG CCAATGTTAT
201 TGTTGTTGGA AACTGTTTTT CATTTGTTAC AAAATAATAC CCTAGAGGAG
251 TTGGCTGGAG GATTTTGAAG GTGCTTCTAT GGCTACTTGC CATTCTCACA
301 GGAATAATAG CTGGCAAATT TCTGTTCCAT CAGCGTTTGT TTGGTCAGTT
351 GCTCCGATTA AAAATGTTTC GAGAAGATCA TGGGTCGTGG ATGACAATGT
401 TCTTCAGCAC AATTCTCTTT CTCTTCATAT TTTCTCACAT ATACAACACG
451 ATTCTTCTAA TGGATGGGAA CATGGGAGCA TATATCATT A CAGACTATAT
501 GGGCATCCGA AATGAAAGTT TCATGAAATT AGCTGCAGTA GGGACCTGGA
551 TGGGGGACTT TGTACAGCT TGGATGGTCA CTGATATGAT GCTTCAGGAC
601 AAACCCATATC CTGACTGGGG AAAATCAGCA AGAGCTTTCT GGAAGAAAGG
651 AAATGTTAGG ATCACTTTAT TCTGGACAGT TCTTTTACT CTGACGCTCG
701 TGGTTGTACT TGTGATTACA ACGGACTGGA TCAGCTGGGA CAAGCTGAAT
751 CGGGGATTTT TGCCCAAGTA TGAAGTTTCC AGAGCATTCC TTGCTTCTTT
801 TATCTTGGTC TTTGACCTTC TTATTGTGAT GCAGGACTGG GAATTCACAC
851 ATTCATGGG AGATGTTGAT GTAAATCTCC CTGGTTTGCA CACCCTCAC
901 ATGCAGTTCA AGATTCTTTT CTTCAGAAA ATCTTCAAGG AGGAATATCG
951 TATTCACATA ACAGGCAAAAT GGTTTAACTA TGAATTATC TTCTCGTCT
1001 TGATTTTGGG TCTTAATATG TGAAGAACC AAATATTTTA TAAACCTCAT
1051 GAATATGGGC AATATATCGG CCCGGGGCAG AAGATATATA CAGTGAAGA
1101 CTCAGAAAGT TTAAAAGATT TGAACAGAAC CAAGCTATCC TGGGAATGGA
1151 GGTCCAATCA CACTAACCTC CGGACTAATA AAACATATGT TGAGGGAGAC
1201 ATGTTCTTAC ACAGCAGGTT CATAGGAGCC AGTCTTGATG TCAAGTGCTT
1251 GGCCTTTGTT CCAAGCCTGA TAGCCTTTGT GTGGTTTGGG TTCTTTATTT
1301 GGTCTTTTGG ACGATTTTGG AAAAATGAGC CACGCATGGA GAATCAAGAC
1351 AAAACTTACA CTCGCATGAA AAGAAAATCT CCATCAGAAC ATAGCAAAGA
1401 CATGGGAATC ACTCGAGAAA ACACCCAGGC TTCAGTAGAA GACCCCTTGA
1451 ATGACCCCTC TTTGGTTTGC ATCAGGCTCG ACTTCAATGA GATCGTCTAC
1501 AAGTCTTCCC ACCTAACCTC GGAAAATTTG AGCTCACAGT TGAACGAATC
1551 TACTAGTGCA ACAGAAGCTG ATCAAGACCC AACGACTTCT AAAAGTACAC
1601 CTACGAACCTA GACTCGGAGA TAGACTTGGA GATAACACAA AAAGCAACCT
1651 TGAGTGTAAC TTTAAAAATT TAGTCTTTCC TTTTGTATAT GTAAGGTTTA
1701 CGTAGTGTTA GGTAAAAATA TGAACAATGC CACAACGGTG CTCACATGCT
1751 TTTTCTTAGG ATTCATTGTT TTCTATTGTT ATTATAATAC ACGTGCCTAC
1801 TGTATACTCA ACAGTCCTCT AGAGATTGCT TTTCACAATT GCACAAGCTA
1851 TTACTGACTT TACAGCATAG TGAAGATTA GCTGATGACC CATGTATCTG
1901 ATGTTCAACC ATAGTGGTGC CTTGAGACAT TAAACTGTTT TTAAGTGTAC
1951 CAGAAATGAA GTGTGGAACA GTTACCTAAC CTATTTCACA TGGGCGTTT
2001 GTATACAACCT ATTTGATCT ACACCTGATG TCTGAGCAGA AAACAGAAAT
2051 AGCTAAATGT GACTCAGGAA GTATCTCTTG GTTCTTATT CAGCAGCAGA
2101 GTTGGTGACT TTGACAACCT GACTGCAGAG AAACATGGTG ATCACCCTTT
2151 AATTTTTATT GGCTGTCTGC CAAATATAAA TACAGATGCA AAATTCAGTA
2201 ATAGGAGATC CATAACCCAA CATGGGTCAC TACTCGTGAA ATGTGACTTT
2251 CTCCACCAG TAATTGAAAT GAGGTGATGA TACCTAATTA TGTTTTCTTA
2301 ATTAAGATA AATTGCTACT TGATTAAAAA TCCTGCCCTT CACCTTTGGG

```

```

2351 AACAAAGGTT AAGAGACACA GTTGGGCGAA CTCTCAAATT TATTGGCATT
2401 TACACAAAGT CCCAGACAAC CAAGGAACGT AAGTTTTCAT CATATGAGAG
2451 CAGCACATCC CACCATTTC CACCATTTAC AATATTCGTA TATCTTCTG CAAATATGGC
2501 TCTGGATAGT GAAAATTGAA AAACATATGC CAACCTGAG CAAGGGAAC
2551 CCTCAAAAAA TCATGCAGCG GAACCTTGTC AGGTAGAGAA GCCGTGCATG
2601 AAAGAATTGT TTTAATGTCT TGTTTTCGCT ATGTGTTTTT TGTTTTTGTT
2651 TTTTAAGAAC TAAATATTGC ACATTAATAA ATAAGAATTA TACAGCAAAA
2701 AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 67 bp to 1608 bp; peptide length: 514
 Category: putative protein

```

1 MGKDFRYFFQ HPWSRMIVAY LVIFFNFIIF AEDPVSHSQT EANVIVVGNC
51 FSFVTNKYPR GVGWRILKVL LWLLAILTGL IAGKFLFHQR LFGQLRLKLM
101 FREDHGSWMT MFFSTILFLF IFSHIYNTIL LMDGNMGAYI ITDYMGIIRNE
151 SFMKLAAVGT WMGDFVTAWM VTDMLQDKP YPDWGKSARA FWKKGNVRIT
201 LEWTVLFTLT SVVVLVITTD WISWDKLNRG FLPSDEVSTRA FLASFILVED
251 LLIVMQDWEF PHFMGDVDVN LPGLHTPHMQ FKIPFFQKIF KEEYRIHITG
301 KWENYGIIFL VLILDNLNWK NQIFYKPEY GQYIGPGQKI YTVKDSESLK
351 DLNRTKLSWE WRSNHTNPR NTYVEGDMF LHSRFIGASL DVKCLAFVPS
401 LIAFVWEGFF IWFFGRFLKN EPRMENQDKT YTRMKRKSPTS EHSKDMGITR
451 ENTQASVEDP LNDPSLVCIR SDFNEIVYKS SHLTSENLSL QLNESTSATE
501 ADQDPPTSKS TPTN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35k24, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35k24, frame 1

Report for DKFZphtes3_35k24.1

```

{LENGTH}      514
{MW}           60185.03
{pI}           8.67
{PROSITE}      MYRISTYL 5
{PROSITE}      CAMP_PHOSPHO_SITE 1
{PROSITE}      CK2_PHOSPHO_SITE 8
{PROSITE}      TYR_PHOSPHO_SITE 1
{PROSITE}      PKC_PHOSPHO_SITE 7
{PROSITE}      ASN_GLYCOSYLATION 6
{KW}           SIGNAL PEPTIDE 32
{KW}           TRANSMEMBRANE 5
{KW}           LOW_COMPLEXITY 15.37 %

```

```

SEQ  MGKDFRYFFQHPWSRMIVAYLVIFFNFIIFAEDPVSHSQT EANVIVVGNCFSFVTNKYPR
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

```

```

SEQ  GVGWRILKVLWLLAILTGLIAGKFLFHQR LFGQLRLKLMFREDHGSWMTMFFSTILFLF
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

```

Prosites for DKFZphtes3_35k24.1

(No Pfam data available for DKFZphtes3_35k24.1)

DKFZphtes3_35n12

group: metabolism

DKFZphtes3_35n12 encodes a novel 315 amino acid protein with strong similarity to ADP/ATP carrier T (ANT) proteins.

The novel protein contains three mitochondrial energy transfer signatures and is closely related to the ADP/ATP translocator, or adenine nucleotide translocator (ANT), a protein most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.

The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

strong similarity to ADP/ATP carrier proteins

EST hits to mouse and drosophila

Sequenced by DKFZ

Locus: unknown

Insert length: 1803 bp

Poly A stretch at pos. 1793, polyadenylation signal at pos. 1772

```
1 AGCGTCCCAA GAGCCACTTT CTCGCCAGTA CGATGCTGCA GCGGTTTTCC
51 GGTTCCTCCG TTCCCTTCAT CGTAGCTCCC GTACTCATTT TTAGCCACTG
101 CTGCCGGTTT TTATATCCTT CTCCATCATG CATCGTGAGC CTGCGAAAAA
151 GAAGGCAGAA AAGCGGCTGT TTGACGCCTC ATCCTTCGGG AAGGACCTTC
201 TGGCCGGCGG AGTCGGGCA GCTGTGTCCA AGACAGCGGT GCGGCCCATC
251 GAGCGGGTGA AGCTGCTGCT GCAGGTGCAG GCGTCGTCGA AGCAGATCAG
301 CCCCAGGCGG CCGTACAAAG GCATGGTGGA CTGCCTGGTG CCGATTCTTC
351 GCGAGCAGGG TTTCTTCAGT TTTTGGCGTG GCAATTTGGC AAATGTTATT
401 CCGTATTTC CAACACAAGC TCTAACTTT GCTTTTAAGG ACAATACAA
451 GCAGCTATTC ATGCTGGAG TTAATAAAGA AAAACAGTTC TGGAGGTGGT
501 TTTTGGCAAA CCTGGCTTCT GGTGGAGCTG CTGGGGCAAC ATCCTTATGT
551 GTAGTATATC CTCTAGATT TGCCCGAACC CGATTAGTG TCGATATTGG
601 AAAAGGTCTT GAGGAGCGAC AATTCAAGGG TTTAGGTGAC TGTATTATGA
651 AAATAGCAAA ATCAGATGGA ATTGCTGGTT TATACCAAGG GTTGGGTGTT
701 TCAGTACAGG GCATCATGTT GTACCGAGCC TCTTATTG GAGCTTATGA
751 CACAGTTAAG GGTATTATC CAAAGCCAAA GAAACTCCA TTTCTTGTCT
801 CCTTTTCAT TGCTCAAGTT GTGACTACAT GCTCTGGAAT ACTTCTTAT
851 CCTTTTGACA CAGTTAGAAG ACGTATGATG ATGCAGAGTG GTGAGGCTAA
901 ACGGCAATAT AAAGGAACCT TAGACTGCTT TGTGAAGATA TACCAACATG
951 AAGGAATCAG TTCCTTTTTT CGTGGCGCCT TCTCCAATGT TCTTCGCGGT
1001 ACAGGGGGTG CTTTGGTGTT GGTATTATAT GATAAAATTA AAGAATTCCT
1051 TCATATTGAT ATTGGTGTA GGTAAATCGG AGAGTAAATT AAGAAATAAC
1101 ATGGATTAA CTTGTAAAC ATACAAATTA CATAGCTGCC ATTTGCATAC
1151 ATTTTGATAG TGTATTGTC TGTATTTGT TAAAGTGCTA GTTCTGCAAT
1201 AAAGCATACA TTTTTCAGG AATTAAATA CTAAATCA GATAAATGTG
1251 GATTTTCCTC CACTTAGAC TCAAACACAT TTTAGTGTGA TATTTCATT
1301 ATTATAGGTA GTATATTTA ATTTGTTAGT TTAATTTCT TTTTATGATT
1351 AAAAATTAAT CATATAATCC TAGATTAATG CTGAAATCTA GGAAATGAAA
1401 GTAGCGTCTT TTAATTCCT ATTCAATTA TATACCTGTT TTCCCATCTT
1451 TTGAAGTCAT ATGGTATGAC ATATTTCTTA AAAGCTTATC AATAGATGTC
1501 ATCATATGTG TAGGCAGAAA TAAGCTTTGT TCTATATCTC TTCTAAGACA
1551 GTTGTATTA CTGTGTATA TATTACAGT ATCAGCCTTT GATTATAGAT
1601 GTGATCATTT AAAATTTGAT AATGACTTTA GTGACATTAT AAACTGAAA
1651 CTGGAATAA AAATGGCTTA TCTGCTGATG TTTATCTTTA AAATAAATAA
1701 AATCTTGCTA GTGTGAATAT ATCTTAGAAC AAAAGGTATC CTCTGAAAAA
1751 TTAGTTTGTA TATTTGTTG ACAATAAAGG AAGCTTAACT GTTAAAAAAA
1801 AAA
```

BLAST Results

No BLAST result

Medline entries

96289608:
Molecular biological and quantitative abnormalities of
ADP/ATP carrier protein in cardiomyopathic hamsters.

Peptide information for frame 2

ORF from 128 bp to 1072 bp; peptide length: 315
 Category: strong similarity to known protein
 Classification: Metabolism
 Prosite motifs: MITOCH_CARRIER (40-50)
 MITOCH_CARRIER (145-155)
 MITOCH_CARRIER (242-252)

```

1 MHREPAKKKA EKRLFDASSE GKDLLAGGVA AAVSKTAVAP IERVKLLQV
51 QASSKQISPE ARYKGMVDCI VRIPREQGFF SFWRGNLAV IRYFPTQALN
101 FAFKDKYKQL FMSGVNKEKQ FWRWFLANLA SGAAGATSL CVVYPLDFAR
151 TRLGVDIGKG PEERQFKGLG DCIMKIAKSD GIAGLYQGFG VSVQGIIVYR
201 ASYFGAYDTV KGLLPKPKKT PFLVSFFIAQ VVTTCSGILS YPFDTVRRRM
251 MMQSGEAKRQ YKGTLDCEVK IQHEGISSF FRGAFSNVLR GTGGALVLVL
301 YDKIKEFFHI DIGGR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n12, frame 2

PIR:S37210 ADP,ATP carrier protein T1 - mouse, N = 1, Score = 1127, P = 2.7e-114

PIR:A44778 ADP,ATP carrier protein T1 - human, N = 1, Score = 1125, P = 4.4e-114

TREMBL:DMADPATPT_2 product: "ADP/ATP translocase"; Drosophila melanogaster gene encoding ADP/ATP translocase, N = 1, Score = 1124, P = 5.6e-114

PIR:XWBO ADP,ATP carrier protein T1 - bovine, N = 1, Score = 1121, P = 1.2e-113

>PIR:S37210 ADP,ATP carrier protein T1 - mouse
 Length = 298

HSPs:

Score = 1127 (169.1 bits), Expect = 2.7e-114, P = 2.7e-114
 Identities = 214/293 (73%), Positives = 248/293 (84%)

```

Query: 17 ASSEFGKDLLAGGVA AAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMVDCI VRIPRE 76
      A SF KD LAGG+AAVSKTAVAPIERVKLLQVQ +SKQIS E +YKG++DC+VRIP+E
Sbjct: 5 ALSFLKDFLAGGIAAAVSKTAVAPIERVKLLQVQHASKQISA EKQYKGIIDCVVRIPKE 64

Query: 77 QGFFSFWRGNLAVIRYFPTQALNFAFKDKYKQLFMSGVNKEKQFWRWFLANLASGGAAG 136
      QCF SFWRGNLAVIRYFPTQALNFAFKDKYKQ+F+ GV++ KQFWR+F NLASGGAAG
Sbjct: 65 QGFLSFWRGNLAVIRYFPTQALNFAFKDKYKQIFLGGVDRHKQFWRYFAGNLASGGAAG 124

Query: 137 ATSLCVVYPLDFARTRLGVDIGKGPEERQFKGLGDCIMKIAKSDGIAGLYQGFGVSVQGI 196
      ATSLC VYPLDFARTRL D+GKG +R+F GLGDC+ KI KSDG+ GLYQGF VSVQGI
Sbjct: 125 ATSLCFVYPLDFARTRLAADVGGSSQREFNGLGDCLTKIFKSDGLKGLYQGFVSVQGI 184

Query: 197 IVYRASYFGAYDTVKGLLPKPKKTPFLVSFFIAQVVTTCSGILSYPFDTVRRRMMMQSGE 256
      I+YRA+YFG YDT KG+LP PK +VS+ IAQ VT +G++SYPFDTVRRRMMMQSG
Sbjct: 185 IIYRAAYFGVYDTAKGMLPDPKNVHII VSWMIAQSVTAVAGLVSYPFDTVRRRMMMQSGR 244

Query: 257 --AKRQYKGTLDCEVKIYQHEGISSFFRGAFSNVLRGTGGALVLVLVYDKIKEF 307
      A Y GTLDC+ KI + EG ++FF+GA+SNVLRG GGA VLVLVD+IK++
Sbjct: 245 KGADIMYTGTDCWRKIAKDEGANAFKGAWSNVLRGMGGAFLVLVLYDEIKKY 297

```

Pedant information for DKFZphtes3_35n12, frame 2

Report for DKFZphtes3_35n12.2

[LENGTH] 315

[illegible]

Prosite for DKFZphtes3_35n12.2

PS00215	40->50	MITOCH_CARRIER	PDOC00189
PS00215	145->155	MITOCH_CARRIER	PDOC00189
PS00215	242->252	MITOCH_CARRIER	PDOC00189

Pfam for DKFZphtes3_35n12.2

HMM_NAME	Mitochondrial carrier proteins		
HMM	*pFwkDFLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpm..ahpRYkGMI +F+KD+LAGG+A++++T+++PI+++K+++Q+Q +++ RYKGM+		
Query	19	SFGKDLLAGGVAAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMV	67
HMM	dCFRwIwkNEGWRGLWRGLgANvIRYIPqWaIRFGFYEFMKemFiDyfge DC+ +I++++G++++WRG++ANVIRY+P++A++F+F++ +K +F + +++		
Query	68	DCLVRIPREQGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLEMSGVVK	117
HMM	ddnyWmWfwmnYMaGsmAGEwisvIitYPMWvVKTRLQaDqkHphsQp.R ++W+WF+ N+++G++AG ++S+ ++YP+++++TRL D +++++ R		
Query	118	EKQFWRWFLANLASGGAAG-ATSLCVVYPLDFARTRLGVD--IGKGPEER	164
HMM	hYNGvWNCwrkIYReEGgFkGLYRGWtPTWMMIPYqmiYFfvYEtLKew +++G+ +C KI +++G ++GLY+G++ +++++I+Y++ YF++Y+T K +		
Query	165	QFKGLGDCIMKIAKSDG-IAGLYQGFGVSVQGIIVYRASYPGAYDTVKGL	213
HMM	lynYtgYnPgprelCMddsPwWhWiIgWmIAGMiaWivSYPFdVVTRMM L +++ + ++++++I++ ++ ++++I+SYPFd+VR+RMM		
Query	214	LP-----KPK--KTPFLVSFFIAQVVT-TCSGILSYPFdTVRRMM	251
HMM	Mdsm.edhkYqSmlDCWMqIYKnEGFkGFWKGFWRPRIMRiMPWtAIMFmI M+S+ ++++Y+++LDC+++IY++EG+ +F++G+ +++R+ ++A+++++		
Query	252	MQSGEAKRQYKGTLDLCFVKIYQHEGISSFFRGAFSNVLRGT-GGALVVLV	300
HMM	YEqMKwFL* Y+ +K+F+		
Query	301	YDKIKEFF	308

DKFZphtes3_35n24

group: testes derived

DKFZphtes3_35n24 encodes a novel 365 amino acid protein without similarity to known proteins.

The novel protein contains a Prosite Ig(Immunoglobulin)-MHC pattern. This pattern represents a domain, approximately one hundred amino acids long and including a conserved intra-domain disulfide bond (Ig domain). Thus, the novel protein is a new member of the Ig-superfamily. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1589 bp

Poly A stretch at pos. 1579, polyadenylation signal at pos. 1560

```

1  CGATCGTCAC GTGACGCCGG GGTTCAGCGT ATCCTTGCTG GGCAACCGTC
51  TTAGAGACCA GCACTGCTGG CTGCACCATG AATGTGATCT ACCCACTGGC
101 AGTCCCCAAG GGGCGCAGAC TCTGCTGTGA GGTGTGCGAA GCGCCAGCCG
151 AGCGGGTGTG CGCGGCCTGC ACAGTCACTT ATTACTGTGG GGTGGTACAT
201 CAGAAGGCTG ACTGGGACAG CATCCATGAG AAAATATGTC AGCTCTTGAT
251 TCCACTGGCG ACTTCCATGC CCTTCTACAA TTCAGAGGAA GAACGGCAGC
301 ATGGCCTGCA GCAGCTGCAG CAGCGGCAGA AGTATTTGAT TGAATTCTGC
351 TACACCATAG CCCAGAAATA CCTCTTTGAA GGGAAACACG AAGATGCTGT
401 ACCAGCAGCT TTGCAGTCCC TTCGCTTCCG TGTGAAGCTG TATGGCCTGA
451 GCTCCGTAGA GCTTGTGCCT GCTTACCCGC TGTGGCCGA GCGCAGCCTT
501 GGTCTGGGCC GAATCGTTCA GGCTGAAGAA TATCTATTCC AAGCCCACTG
551 GACAGTCCCT AAATCAACTG ACTGTAGTAA TGCCACCCAC TCTTTACTGC
601 ATCGGAATCT GGGACTTCTC TATATAGCTA AGAAAACTA TGAAGAGGCC
651 CGTTATCATC TGGCCAATGA TATTTATTTT GCCAGTTGTG CATTTGGAAC
701 AGAGGACATT AGGACTTCAG GAGGCTACTT CCACCTGGCT AATATATTCT
751 ATGACCTTAA AAAGTTGGAC CTGGCAGACA CATTGTACAC CAAGGTCTCT
801 GAGATCTGGC ATGCATATTT GAACAATCAC TATCAAGTCC TCTCACAGGC
851 TCACATCCAA CAAATGGATT TACTGGGCAA ACTATTGAG AATGACACTG
901 GCTTGGATGA AGCCCAAGAA GCAGAAGCCA TTCGCATCCT GACTTCAATC
951 TTGAACATTC GAGAATCTAC ATCTGACAAA GCCCCCCAAA AAACCATCTT
1001 TGTTCTGAAG ATCCTGGTCA TGCTTTACTA CCTGATGATG AATTCTTCAA
1051 AGGCACAGGA ATATGGCATG AGGGCCCTCA GTCTAGCCAA AGAACAACAG
1101 CTTGATGTCC ATGAGCAAAG CACCATTCAA GAGTTATTAA GTCTCATTTT
1151 AACTGAAGAC CATCCCATTA CTTAGTGACC CATGAGCTCT GCATCAAGGG
1201 TTATTCCAGG GGCTACTGAA GATCTAATAT ATTCCAGCCT TGCACAACCTG
1251 CTTTGAGGTA CTGTAGACTG CTGAAGTTTC CACCCTCTTC CCCTGGGATT
1301 GCACACATAG CTGTTATTTT TTTCTTACAC AGCATATTAA GGGAAATATA
1351 AGCTTTAGGC ATAGAAATCA CTAAAACTG TGTTTGTGTC GACCTTTGTA
1401 CTTGATTTAT CATGACTTTG TATGACTGAG TAATATGTAG TCAGATCACT
1451 AATATGGTAT TTGTAATTAA ACTACAAATA GTTTGTCATT TCCCAGAAGT
1501 CTTCCAACGA TGCATGTTTC ATACACTTTT GCTAAAGGAG GGGTAAAGGA
1551 GGGGGTAGGG AATAAAGCTA TATTGGAACA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 78 bp to 1172 bp; peptide length: 365
Category: putative protein

Prosites motifs: IG_MHC (35-42)

```

1 MNVIYPLAVP KGRRLCCEVC EPAERVCAA CTVTYCGVV HQKADWDSIH
51 EKICQLLIPL RTSMPFYNSE EERQHGLQQL QORQKYLIEF CYTIAQKYL
101 EKGHEDAVPA ALQSLRFRVK LYGLSSVELV PAYPLAEAS LGLGRIVQAE
151 EYLFQAQWTV LKSTDCSNAT HSLLRNLGL LYIAKKNYEE ARYHLANDIY
201 FASCAFGTED IRTSGGYFHL ANIFYDLKKL DLADTLTKV SEIWHAYLNN
251 HYQVLSQAHQ QMDLLGKLF ENDTGLDEAQ EAEAIRILTS ILNIRESTSD
301 KAPQKTIFVL KILVMLYYLM MNSSKAQYEG MRALSLAKEQ QLDVHEQSTI
351 QELLSLISTE DHPIT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n24, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35n24, frame 3

Report for DKFZphtes3_35n24.3

```

[LENGTH]      365
[MW]           41768.24
[pI]           5.82
[BLOCKS]      BL00273 Heat-stable enterotoxins proteins
[PROSITE]     MYRISTYL 1
[PROSITE]     IG_MHC 1
[PROSITE]     AMIDATION 1
[PROSITE]     CK2_PHOSPHO_SITE 7
[PROSITE]     TYR_PHOSPHO_SITE 4
[PROSITE]     PKC_PHOSPHO_SITE 3
[PROSITE]     ASN_GLYCOSYLATION 3
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY 4.11 %

```

```

SEQ  MNVIYPLAVPKGRRLCCEVCEAPAERVCAACTVTYTCGVVHQKADWDSIHEKICQLLIPL
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  RTSMPFYNSEEERQHGLQQLQORQKYLIEFCYTIAQKYLFEKGHEDAVPAALQSLRFRVK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LYGLSSVELVPAYPLAEASLGLGRIVQAEYLFQAQWTVLKSTDCSNATHSLLRNLGL
SEG  .....
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LYIAKKNYEEARYHLANDIYFASCAFGTEDIRTSGGYFHLANIFYDLKKLDLADTLTKV
SEG  .....
PRD  eeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  SEIWHAYLNNHYQVLSQAHIQMDLLGKLFENDTGLDEAQEAEAIRILTSILNIRESTSD
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KAPQKTIFVLKILVMLYYLMNNSKAQYEGMRALSLAKEQQLDVHEQSTIQELLSLISTE
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  DHPIT
SEG  .....
PRD  ccccc

```

Prosites for DKFZphtes3_35n24.3

PS00001	168->172	ASN_GLYCOSYLATION	PDOC00001
PS00001	272->276	ASN_GLYCOSYLATION	PDOC00001
PS00001	322->326	ASN_GLYCOSYLATION	PDOC00001
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005

PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	69->73	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00007	85->93	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	185->194	TYR_PHOSPHO_SITE	PDOC00007
PS00008	275->281	MYRISTYL	PDOC00008
PS00009	11->15	AMIDATION	PDOC00009
PS00290	35->42	IG_MHC	PDOC00262

(No Pfam data available for DKFZphtes3_35n24.3)

DKFZphtes3_35n9

group: metabolism

DKFZphtes3_35n9 encodes a novel 607 amino acid protein which is a splice variant of human carboxylesterase (EC 3.1.1.1).

The novel protein contains both, one carboxylesterase B1 and one B2 pattern. In comparison to EC 3.1.1.1, DKFZphtes3_35n9 shows a N-terminal extension and aa 458-474 are missing.

The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

carboxylesterase, splice variant

5' extension of mRNA and N-terminal elongation of protein (64 aa),
missing exon! aa 458-474 of JC5408 are missing

Sequenced by DKFZ

Locus: unknown

Insert length: 2888 bp

Poly A stretch at pos. 2878, no polyadenylation signal found

```

1 CTCGGCCTGA GGTGCGAGAG AAGCGGTGAC CGCGGCCCTG GCTGCTCGGA
51 CCCGGGAACA TGATGGTCCG TGGAGCAGAA GGCCTGAGA AGGGACCACG
101 GCGGCGCTGG GTCGTGCGAG CCAGTAGCGG GCTGAAACGT AGAGGCCAGA
151 ACCAGGTCTC AGGGGGCACT AAAGCGCGTC GGAGGTAATC CCCACACCGC
201 TTCTCTCTGG AAGTCAGGCT GGCCGGGAGC TCCCGTATCC AGGACGGTTG
251 GTCGCCCTCTG GCCTGGCAGG GATCCTAGTG TCTCGGGACC TCCCGGTGAC
301 GCGCCTGCCT CCCCTGCTGC ACCATAGGCC CGGGAGTACG GCGTCCCCAC
351 AGCTTGGACC GGCAGGGGCT CGTGAAATGT TTGTCAAGTG GATAAATGAC
401 CATGGCCGTG GTCTCCGCGG GAGGTGAGGA AACTGAAAGC CACCGAGGAA
451 AAGGGGGGCG CTCCTTAAGA AGTGCCGCGG TCACGTGTAC GTTTCAAAAG
501 AATGGCGTGA CTGAGTAGGG AGGGGACCGC GGAGACCCCT AGACCCTGGA
551 CTGTAAGGAG ATGAGGGGCC GTGAAGGGGA ACCCAGGAAA CTGAGTCCTG
601 AAAGCAAGGA GGAACCTTCCA GAATGAAGGG CGCCGACACT CCTTCTGCCC
651 TTTGCTCAAG CGGTTCCCTC ACCCCGATCA AGTTCCTTCC CATTCTCTCA
701 TCTGGGGGAT CCTGAACGTG CACATCCTCA GAGAAGCCCT CCTGGGGTCT
751 CCAATTCATG TTTATTGCCC CCTCTATCG ATCCCCAGC GCGCTCATCG
801 GGCCTGTGGA CAAGGACAGG TTTGAAGAGA GGATTCCCTG GATCGCGGAA
851 GGGCTGCAGG AATGGCACAG CCCCTTCCGA GGATGCCAAA GGAGCCCGGG
901 CAAAGGAAAG TGGCCGTGCC CGGGCCTGCC TACCACTAGA TCCCCACCCA
951 CCTATGACTG CTCAGTCCCG CTCTCCTACC ACACCCACCT TCCCCGGCCC
1001 AAGCCAGCGC ACCCCGCTGA CTCCCTGCCC AGTCCAAACT CCAAGGCTGG
1051 GCAAGGCACT GATCCACTGC TGGACAGACC CGGGGACGCC TCTGGGTGAA
1101 CAGCAGCGTG TCCGCCGGCA GCGAACCAGG ACCAGCGAGC CGACCATGCG
1151 GCTGCACAGA CTTCTGTGCG GGCTGAGCGC GGTGGCCTGT GGGCTTCTGC
1201 TGCTTCTTGT CCGGGGCCAG GGCCAGGACT CAGCCAGTCC CATCCGGACC
1251 ACACACACGG GGCAGGTGCT GGGGAGTCTT GTCCATGTGA AGGGCGCCAA
1301 TGCCGGGGTC CAAACCTTCC TGGGAATTCC ATTTGCCAAG CCACCTCTAG
1351 GTCCGCTGCG ATTTGCACCC CCTGAGCCCC CTGAATCTTG GAGTGGTGTG
1401 AGGGATGGAA CCACCCATCC GGCCATGTGT CTACAGGACC TCACCCGAGT
1451 GGAGTCAGAG TTTCTTAGCC AGTTCAACAT GACCTTCCCT TCCGACTCCA
1501 TGTCGAGGA CTGCCTGTAC CTCAGCATCT ACACGCCGGC CCATAGCCAT
1551 GAAGGCTCTA ACCTGCCGGT GATGGTGTGG ATCCACGGTG GTGCGCTTGT
1601 TTTTGGCATG GCTTCTTGT ATGATGGTTC CATGCTGGCT GCCTTGAGGA
1651 ACGTGGTGGT GGTCAATCAT CAGTACCGCC TGGGTGTCTT GGGCTTCTTC
1701 AGCACTGGAG ACAAGCACGC AACCGGCAAC TGGGGCTACC TGGACCAAGT
1751 GGCTGCACTA CGCTGGGTCC AGCAGAATAT CGCCCACTTT GGAGGCAACC
1801 CTGACCGTGT CACCATTTTT GGCGAGTCTG CGGGTGGCAC GAGTGTGTCT
1851 TCGCTTGTGT TGTCCCCCAT ATCCCAAGGA CTCTTCCACG GAGCCATCAT
1901 GGAGAGTGGC GTGGCCCTCC TGCCCGGCCT CATTGCCAGC TCAGCTGATG
1951 TCATCTCCAC GGTGGTGGCC AACCTGTCTG CCTGTGACCA AGTTGACTCT
2001 GAGGCCCTGG TGGGCTGCCT GCGGGGCAAG AGTAAAGAGG AGATTCTTGC
2051 AATTAACAAG CCTTTCAAGA TGATCCCCGG AGTGTGGAT GGGGTCTTCC
2101 TGCCCAGGCA CCCCAGGAG CTGCTGGCCT CTGCCGACTT TCAGCCTGTC
2151 CCTAGCATTG TTGGTGTCAA CAACATGAA TTCGGCTGGC TCATCCCCAA
2201 GGTCATGAGG ATCTATGATA CCCAGAAGGA AATGGACAGA GAGGCTTCCC
2251 AGGCTGTCTC GCAGAAAAATG TTAACGCTGC TGATGTTGCC TCCTACATTT
2301 GGTGACCTGC TGAGGGAGGA GTACATTGGG GACAATGGGG ATCCCCAGAC
2351 CCTCCAAGCG CAGTTCCAGG AGATGATGGC GGACTCCATG TTTGTGATCC
2401 CTGCACTCCA AGTAGCACAT TTTAGTGTTC CCCGGGCCCT TGTGTACTTC
2451 TACGAGTTCC AGCATCAGCC CAGCTGGCTC AAGAACATCA GGCCACCGCA
2501 CATGAAGGCA GACCATGTTA AATTCATGTA GGAAGAGGAG CAGCTAAGCA
2551 GGAAGATGAT GAAGTACTGG GCCAATTTG CGAGAATGG GAACCCCAAT
2601 GGCGAGGGTC TGCCACACTG GCCGCTGTTT GACCAGGAGG AGCAATACCT

```

2651 GCAGCTGAAC CTACAGCCTG CGGTGGGCCG GGCTCTGAAG GCCCACAGGC
 2701 TCCAGTTCTG GAAGAAGGCG CTGCCCCAAA AGATCCAGGA GCTCGAGGAG
 2751 CTTGAAGAGA GACACACAGA GCTGTAGCTC CCTGTGCCGG GGAGGAGGGG
 2801 GTGGGTTTCG TGACAGGCGA GGGTCAGCCT GCTGTGCCCA CACACACCCA
 2851 CTAAGGAGAA AGAAGTTGAT TCCTTCATAA AAAAAAAA

BLAST Results

Entry D50579 from database EMBL:
 Homo sapiens mRNA for carboxylesterase, complete cds.
 Score = 7197, P = 0.0e+00, identities = 1441/1443

Entry JC5408 from database PIR:
 carboxylesterase (EC 3.1.1.1) - human
 Score = 2808, P = 1.2e-291, identities = 542/559, positives = 543/559,
 frame +3

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 954 bp to 2774 bp; peptide length: 607
 Category: known protein
 Classification: Metabolism
 Prosite motifs: CARBOXYLESTERASE_B_1 (279-295)
 CARBOXYLESTERASE_B_2 (185-196)

1 MTAQSRSPPT PTFPGPSQRT PLTPCPVQTP RLKALIHCV TDPGQPLGEQ
 51 QRVRRQRIT SEPTMRLHRL RARLSAVACG LLLLLVRGQG QDSASPIRTT
 101 HTGQVLGSLV HVKGANAGVQ TFLGIPFAKP PLGPLRFAPP EPPESWSGVR
 151 DGTTHPAMCL QDLTAVESEF LSQFNMTFPS DSMSEDCLYL SIYTPAHSHE
 201 GSNLPVMVWI HGGALVFGMA SLYDGSMLAA LENVVVVVIQ YRLGVLGFFS
 251 TGDKHATGNW GYLDQVAALR WVQQNIAHEG GNPDRVITFG ESAGGTSVSS
 301 LVVSPISQGL FHGAIMESGV ALLPGLIASS ADVISTVVAN LSACDQVDSE
 351 ALVGCLRGKS KEEILAINKP FKMIPGVVDG VFLPRHPQEL LASADFQVPV
 401 SIVGVNNNEF GWLIPKVMRI YDTQKEMDRE ASQAALQKML TLLMLPPTFG
 451 DLLREEYIGD NGDPQTLOAQ FQEMMADSMF VIPALQVAHF QCSRAPVYFY
 501 EFQHQPSWLK NIRPPHMKAD HVKFTSEEEQ LSRKMMKYWA NFARNGNPNP
 551 EGLPHWPLFD QEEQYLQLNL QPAVGRALKA HRLQFWKKAL PQKIQELEEP
 601 EERHTEL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n9, frame 3

PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human, N = 1, Score = 2808,
 P = 1.9e-292

TREMBL:HSU60553_1 gene: "hCE-2"; product: "carboxylesterase"; Human
 carboxylesterase (hCE-2) mRNA, complete cds., N = 1, Score = 2761, P =
 1.8e-287

PIR:A34329 60K esterase (EC 3.1.1.-) isoform 2 - rabbit, N = 1, Score =
 1985, P = 3.1e-205

TREMBL:D50580_1 product: "carboxylesterase precursor"; Rattus
 norvegicus mRNA for carboxylesterase, partial cds., N = 1, Score =
 1984, P = 4e-205

>PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human
 Length = 559

HSPs:

Score = 2808 (421.3 bits), Expect = 1.9e-292, P = 1.9e-292

Identities = 542/559 (96%), Positives = 543/559 (97%)

Query: 65 MRLHRLRLRLSAVACGLLLLLLVGGQGDSPARTHTTQVGLSLVHVKGANAGVQTFGLG 124
 MRLHRLRLRLSAVACGLLLLLLVGGQGDSPARTHTTQVGLSLVHVKGANAGVQTFGLG
 Sbjct: 1 MRLHRLRLRLSAVACGLLLLLLVGGQGDSPARTHTTQVGLSLVHVKGANAGVQTFGLG 60

Query: 125 IPFAKPLPLRFAPPEPPESWSGVRDGTTHPAMCLODLTAVESEFLSQFNMTFFSDSMS 184
 IPFAKPLPLRFAPPEPPESWSGVRDGTTHPAMCLODLTAVESEFLSQFNMTFFSDSMS
 Sbjct: 61 IPFAKPLPLRFAPPEPPESWSGVRDGTTHPAMCLODLTAVESEFLSQFNMTFFSDSMS 120

Query: 185 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVIIQYRLG 244
 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVIIQYRLG
 Sbjct: 121 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVIIQYRLG 180

Query: 245 VLGFSTGDKHATGNWGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS 304
 VLGFSTGDKHATGNWGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS
 Sbjct: 181 VLGFSTGDKHATGNWGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS 240

Query: 305 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEI 364
 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEI
 Sbjct: 241 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEI 300

Query: 365 LAINKPFKMI PGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ 424
 LAINKPFKMI PGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ
 Sbjct: 301 LAINKPFKMI PGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ 360

Query: 425 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA 484
 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA
 Sbjct: 361 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA 420

Query: 485 LQVAHFQCSRAPVYFYEFGHQPSWLKNIRPPHMKADH-----VKFTEEE 528
 LQVAHFQCSRAPVYFYEFGHQPSWLKNIRPPHMKADH +KFTEEE
 Sbjct: 421 LQVAHFQCSRAPVYFYEFGHQPSWLKNIRPPHMKADHGDELPEVFRSFFGGNYIKFTEEE 480

Query: 529 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 588
 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK
 Sbjct: 481 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 540

Query: 589 ALPQKIQELEEEPEERHTEL 607
 ALPQKIQELEEEPEERHTEL
 Sbjct: 541 ALPQKIQELEEEPEERHTEL 559

Pedant information for DKFZphtes3_35n9, frame 3

Report for DKFZphtes3_35n9.3

[LENGTH] 607
 [MW] 67051.20
 [pI] 6.11
 [HOMOL] PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human 0.0
 [BLOCKS] BL01173A Lipolytic enzymes "G-D-X-G" family, histidine
 [BLOCKS] BL00122G
 [BLOCKS] BL00122F
 [BLOCKS] BL00122E
 [BLOCKS] BL00122D Carboxylesterases type-B serine proteins
 [BLOCKS] BL00122C Carboxylesterases type-B serine proteins
 [BLOCKS] BL00122B Carboxylesterases type-B serine proteins
 [BLOCKS] BL00122A Carboxylesterases type-B serine proteins
 [SCOP] dlakn_ 3.56.1.1.4 Bile-salt activated lipase [Bovine (Bos taurus 1e-158
 [SCOP] d2ack_ 3.56.1.1.1 Acetylcholinesterase [Electric ray (Torped 1e-170
 [SCOP] dlthg_ 3.56.1.9.7 type-B carboxylesterase/lipase [fungu 1e-149
 [EC] 3.1.1.13 Sterol esterase 1e-52
 [EC] 3.1.1.7 Acetylcholinesterase 5e-74
 [EC] 3.1.1.1 Carboxylesterase 0.0
 [EC] 3.1.1.8 Cholinesterase 5e-68
 [EC] 3.1.1.59 Juvenile-hormone esterase 1e-34
 [EC] 3.1.1.3 Triacylglycerol lipase 3e-52
 [PIRKW] duplication 2e-47
 [PIRKW] homotetramer 3e-67
 [PIRKW] transmembrane protein 9e-44
 [PIRKW] microsome 1e-130
 [PIRKW] pancreas 3e-52
 [PIRKW] endoplasmic reticulum 1e-134
 [PIRKW] homotrimer 1e-134
 [PIRKW] phosphatidylinositol linkage 5e-74
 [PIRKW] synapse 3e-73
 [PIRKW] liver 1e-131
 [PIRKW] heparin binding 3e-52

```

[PIRKW]      phosphoprotein 7e-25
[PIRKW]      glycoprotein 1e-134
[PIRKW]      thyroid hormone biosynthesis 2e-47
[PIRKW]      carboxylic ester hydrolase 0.0
[PIRKW]      monomer 2e-42
[PIRKW]      disulfide bond 2e-31
[PIRKW]      mammary gland 3e-52
[PIRKW]      alternative splicing 5e-74
[PIRKW]      iodine 2e-47
[PIRKW]      pyroglutamic acid 6e-39
[PIRKW]      hydrolase 1e-135
[PIRKW]      muscle 3e-73
[PIRKW]      thyroid gland 2e-47
[PIRKW]      membrane protein 3e-73
[PIRKW]      neurotransmitter degradation 3e-73
[PIRKW]      cholesterol 3e-52
[PIRKW]      homodimer 2e-47
[PIRKW]      nerve 3e-73
[SUPFAM]     cholinesterase 0.0
[SUPFAM]     triacylglycerol lipase 1e-32
[SUPFAM]     cholinesterase homology 0.0
[SUPFAM]     thyroglobulin 2e-47
[SUPFAM]     thyroglobulin type I repeat homology 2e-47
[SUPFAM]     juvenile-hormone esterase 2e-35
[SUPFAM]     probable lipolytic protein ybaC 1e-07
[PROSITE]    CARBOXYLESTERASE_B_2 1
[PROSITE]    CARBOXYLESTERASE_B_1 1
[PFAM]       Carboxylesterases
[KW]         Alpha_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY      3.95 %

```

```

SEQ      MTAQSRSPPTPTFPGPSQRTPLTPCPVQTPRLGKALHCWTDPGQPLGEQQRVRRQRTET
SEG      .....XXXXXXXXX...
lacj-    .....

SEQ      SEPTMRLHRLRLRLSAVACGLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQ
SEG      .....XXXXX.....
lacj-    .....ETEEEECEEEEEETEE--EE

SEQ      TFLGIPFAKPLPLRFAPPEPPESWSGVVDGTTTHPAMCLQDLTAVESEFLSQFNMTFFS
SEG      .....
lacj-    EEEEECEETTTGGGTTTCCECCCCCEEECCCCBCCCCCTTTTT-HHHHHCCCC

SEQ      DSMSEDCLYLSIYTPAHSHEGSLNLPVMVWINGGALVFGMASLYDGSMLAALENVTVVVIQ
SEG      .....
lacj-    CCBTTTTCEEEEEET--TTTTTEEEEEECTTTTTCTTTTGCHHHHHHHHCCEEEEC

SEQ      YRLGVLGFSTGDKHATGNWGYLDQVAALRWVQQNIAHEFGNPDRTVIFGESAGGTSVSS
SEG      .....
lacj-    CCCCCGGGCTTTTTTTCCHHHHHHHHHHHHHHCGGGGCEEEEEEECHHHHHHHH

SEQ      LVVSPISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKS
SEG      .....
lacj-    HHHCGGGTTTTCEEEETTTTTTTTTTCHHHHHHHHHHHHC-CCCCCHHHHHHHHHHCC

SEQ      KEEILAINKPKFMI PGVVDGVFLPRHPQELLASADFQVPVSI VGVNNNEFGWLI PKVMRI
SEG      .....
lacj-    HHHHHHHHTCCCTTTCTTTTTTTTTTHHHHHHTTCCCEEEEEETBTHHHHHHTTTT

SEQ      YDTQKEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMF
SEG      .....
lacj-    TTTCCCCCHHHHHHHHHHTTTTCHHHHHHHHHHCTTTTTTHHHH-HHHHHHHHHHHH

SEQ      VIPALQVAHFQCSRAPVYFYEFQHQPSWLKNIRPPHMKADHVKFTEEEELSRKMMKYWA
SEG      .....
lacj-    HHHHHHHHHHHHCCCCEEEEECCECGGTTBTTHHHCGGGCCCHHHHHHHHHHHHHH

SEQ      NFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKKALPQKIQEELEP
SEG      .....XXXXX
lacj-    HHHHHCCCCCCC--CCCCBTBTBEEECCECEETTHHHHHHHHHHHH.....

SEQ      EERHTEL
SEG      xxxxxx.
lacj-    .....

```

Prosites for DKFZphtes3_35n9.3

PS00122 279->295 CARBOXYLESTERASE_B_1 PDOC00112
 PS00941 185->196 CARBOXYLESTERASE_B_2 PDOC00112

Pfam for DKFZphtes3_35n9.3

HMM_NAME	Carboxylesterases
HMM	*MfMnwlimFLLwmItWii.WheqaprpPdPyivdtnnCGkIRGmNedtD + +L+++ ++++++ ++Q+++P I T+ G + G ++ +
Query	69 RLRARLSAVACGLLLLVRGQGQDSASP---IRTTHT-GQVLGSLVHVK 113
HMM	NG..pYYvFlGIPYAEPVGNLRFKePQPYhePWtNVWNATnYPPMCMQW + + +FLGIP+A+PP+G LRF +P+P +E W++V++ T+ P MC+Q+
Query	114 GANAGVQTFLGIPFAKPLGLRFAPEP-PESWSGVRDGTTHPAMCLQD 162
HMM	ndFGFWlFdmieMWNeniP..eMSEDCLYLNVTWPNrkPNskLPVMVWI +++ ++N++ P +MSEDCLYL+++TP+ + ++S+LPVMVWI
Query	163 LTAV--ESEFLSQFNMTFPSDSMSEDCLYLSIYTPAHSHEGSNLPVMVWI 210
HMM	HGGGFMEFGSGhsYPLiqYDgeylMMeenVIVvtINyRLGPFGLSTgDid HGG+++FG + ++YDG+ L++ ENV+VV I+YRLG++GF+STGD +
Query	211 HGGALVFGMA-----SLYDGSMLAALENVVVVVIQYRLGVLFSTGDKH 255
HMM	lPPHGNWGLWQRMALQWVDNianFGGDPNNITIFGESAGGMSVHlHML + GNWG++DQ++AL+WVQ+NIA+FGG+P+++TIFGESAGG+SV+ ++
Query	256 AT--GNWGYLDQVAALRWVQNIaHEGNGNPDRTIFGESAGGTSVSSLV 303
HMM	SYGGDNPPmfKqLFHRAIMQSGsAmcPWvIQsnyNaRqRAfRFArimGCN S P + +LFH AIM+SG A+ P++I S++ + +A++ C+
Query	304 S-----PISQGLFHGAIMESGVALPLGLIASSA--DVISTVVANLSACD 345
HMM	rmDsseMIqCLRSKPwEELWdAtWnFWmWfYfPflPWFFgPVIDGDDaPE + DS++++ CLR K+ EE+++++ +F + + +DG+
Query	346 QVDSEALVGCLRGKSKEELAINK----PFKMIPGV-----VDGV---- 381
HMM	aFIPDHPeemIKEGkFnDVPWIIIGYnNDEGiWFapMmMnfnWfdEDeWId F+P+HP+E++++ F VP I+G+NN E++W++P M + + +E++
Query	382 -FLPRHPQELLASADFQPVFSIVGVNNNEFGWLIPKVMRIYDT-QKEMDR 429
HMM	itNedWyeWMPYIlFYrddmsNikDMDDYiDkvyEeYPgWWDrfPqESYW ++ + ++ M +L + + + D ++EEY+G+ + PQ
Query	430 EASQAALQKMLTLMLPPT-F-----GDLLREEYIGDNGD-PQTLQA 469
HMM	nLqDMFTDYLFWCpRihadnHRkHwgsPVYMYeFDHPPsFGYgQFFmWR ++Q+M+ D F++P + ++H++ +PVY+YEF+H PS +
Query	470 QFQEMMADSMFVIP--ALQVAHFQCSRAPVYFYEFQHQPSW-----LKN 511
HMM	WWPpWMgvDH* +PP+M++DH
Query	512 IRPPHMKADH 521
HMM	*tEEEiissMRmMMNYWINFaKhGNPNnthnglCWWPqYTsneQYdMIME TEEE+ +S R MM+YW+NFA++GNPN++ GL++WP ++++EQY++ +
Query	525 TEEEEEQLS-RKMMKYWANFARNGNPNGE--GLPHWPLFDQEEQYLQNL 570
HMM	tIImiQmCrmrDPYCNFW* + +++++ + FW
Query	571 QPAVGRALKAHR--LQFW 586

DKFZphtes3_35p17

group: testes derived

DKFZphtes3_35p17 encodes a novel 505 amino acid protein with weak similarity to proteins of the armadillo family.

Proteins of the armadillo family are involved in diverse cellular processes in higher eukaryotes. Some of them, like armadillo, beta-catenin and plakoglobins have dual functions in intercellular junctions and signalling cascades. Others, belonging to the importin-alpha-subfamily are involved in NLS recognition and nuclear transport, while some members of the armadillo family have as yet unknown functions. The novel protein shows similarity to *S. cerevisiae* protein Yel013p (VAC8) and *Danio rerio* b-catenin, but contains no armadillo (arm) repeats.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S. cerevisiae* VAC8

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1966 bp

Poly A stretch at pos. 1956, polyadenylation signal at pos. 1935

```
1 AAGTCAAATG TAAGATTGGT TCATTAAAAA TACTGAAGGA AATCAGTCAT
51 AATCCTCAAA TCAGACAGAA TATTGTTGAC CTGGGGGCT TACCAATTAT
101 GGTGAATATA CTTGATTCTC CACACAAGAG TCTAAAATGT TTGGCAGCCG
151 AGACTATCGC GAATGTTGCC AAGTTTAAAA GAGCACGGCG GGTGGTGAGG
201 CAGCACGGGG GTATCACCAA ACTGGTTGCT CTACTAGACT GTGCACATGA
251 TTCCACAAAA CCTGCCCAAT CGAGTCTGTA TGAGGCCAGA GACGTGGAAG
301 TGGCTCGCTG TGGGGCACTG GCCCTGTGGA GCTGCAGTAA GAGTCATACG
351 AATAAAGAAG CCATCCGCAA AGCTGGGGGC ATTCTCTGT TGGCTCGGCT
401 GCTGAAGACT TCTCATGAAA ACATGCTAAT TCCAGTGGTG GGGACATTGC
451 AAGAGTGTGC ATCAGAGGAA AACTACCGGG CTGCAATCAA AGCAGAAAGG
501 ATCATTGAAA ACCTTGTCAA GAACCTAAAT AGTGAGAATG AGCAGCTGCA
551 GGAGCACTGC GCCATGGCCA TTTACCAGTG TGCTGAAGAT AAGGAAACCC
601 GGGACCTCGT TAGGCTGCAC GGAGGACTTA AGCCCTTGGC CAGTCTACTC
651 AATAACACTG ACAATAAAGA GCGGTTAGCT GCTGTCACAG GGGCTATATG
701 GAAATGTTCC ATCAGCAAAG AGAATGTTAC CAAGTTTCGG GAATACAAAG
751 CCATTGAAAC CTGGTGGA CTCTAACAG ATCAGCCTGA AGAAGTACTT
801 GTGAATGTGG TTGGGGCCTT GGGAGAATGC TGCCAAGAAC GTGAAAACCG
851 AGTCATTGTC CGGAAATGTG GTGGCATTCA ACCACTTGTG AACCTCCTTG
901 TTGGAATAAA CCAAGCTCTT CTTGTGAATG TTACAAAAGC AGTTGGTGCT
951 TGTGCAGTAG AACCTGAAAAG TATGATGATA ATTGATCGCT TAGATGGAGT
1001 TCGTTTGTG TGGTCCCTGC TGA AAAATCC TCACCCAGAC GTGAAGGCCA
1051 GCGCAGCATG GGCACCTGTG CCATGCATCA AAAATGCAAA GGATGCTGGG
1101 GAAATGGTTC GTTCCCTTGT TGGTGGTTTG GAACCTATTG TCAATTTACT
1151 GAAATCAGAT AACAAAGAAG TTCTGGCAAG TGTATGTGCT GCCATTACCA
1201 ACATAGCAAA AGATCAAGAA AATTAGCTG TTATCACAGA TCATGGAGTT
1251 GTTCCCTTAT TGTCCTAACT GGCAATACA AATAACAATA AATTGAGACA
1301 TCATCTAGCA GAAGCTATTT CACGTTGCTG TATGTGGGGC AGGAATAGAG
1351 TGGCCTTCGG TGAGCACAAA GCAGTGGCTC CACTAGTGGC TTATCTGAAA
1401 TCAAATGACA CCAACGTGCA TCGGGCGACA GCTCAGGCCT TGTACCAACT
1451 CTCAGAAGAC GCCGATAACT GCATCACCAT GCATGAGAA TGGTGCAGTAA
1501 AGCTTCTACT GGATATGGTT GGGTCCCTG ACCAGGATCT CCAGGAAGCT
1551 GCAGCTGGTT GTATATCCAA TATCCGAGG CTGGCTCTTG CTACAGAGAA
1601 GGCAAGATAC ACTTGAAATT TAAATGGACA TTACAAGCTA TCAAATTCTA
1651 CATGACACAG GACATGTCAC TCCATGGGCC AGAAAGCCTA AATTGGGAAA
1701 CAGTTGTTAG CAACCCCTTT CAACCATCTA AATGAAAACA CACAAATTGA
1751 AATGACACAG AATGTTTTTC ATCTGAAAAT TGATGGAGA CTTTGTGTTT
1801 TATTTAATGT TTTGAGATA TGACATGTGA TAAGATGGAA AGCCAATAAA
1851 CCGTGATATA GTTCTAAGA ATATGAGAA ATACGTATAT GATGATTTT
1901 TAGTTCAGTG ATGCTTTTGT ATTTGTGGCG ATTTTAATAA AGGATATGGC
1951 CTTCCTCAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

98413148:
Yel013p (Vac8p), an armadillo repeat protein related to plakoglobin and importin alpha is associated with the yeast vacuole membrane.

98330438:
YEB3/VAC8 encodes a myristylated armadillo protein of the *Saccharomyces cerevisiae* vacuolar membrane that functions in vacuole fusion and inheritance.

98158703:
Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole.

Peptide information for frame 3

ORF from 99 bp to 1613 bp; peptide length: 505
Category: similarity to known protein
Classification: unset

```

1 MVNILDSPHK SLKCLAAETI ANVAKFKRAR RVVRQHGGIT KLVALLDCAH
51 DSTKPAQSSL YEARDVEVAR CGALALWSCS KSHTNKEAIR KAGGIPLLAR
101 LLKTSHEMML IPVVGTLQEC ASEENYRAAI KAERIIENLV KNLNSENELQ
151 QEHCAAIYQ CAEDKETRD LRLHGGGLKPL ASLLNNTDNK ERLAAVTGAI
201 WKCSISKENV TKFREYKAIE TLVGLLTDQP EEVLNVVGA LGCECCQEREN
251 RVIVRKCGGI OPLVNLLVGI NQALLVNVTK AVGACAVEPE SMMIIDRLDG
301 VRLWLSLLKN PHPDVKASAA WALCPCIKNA KDAGEMVRSF VGGLELIVNL
351 LKSDNKEVLA SVCAAITNIA KDQENLAVIT DHGVVPLLSK LANTNNKLR
401 HHLAEAISRC CMWGRNRVAF GEHKAVAPLV RYLSNDTNV HRATAQALYQ
451 LSEDADNCIT MHENGAVKLL LDMVGSPDQD LQEAAGCIS NIRRLALATE
501 KARYT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p17, frame 3

PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 237, P = 7.8e-17

PIR:T00403 T13E15.9 protein - *Arabidopsis thaliana*, N = 1, Score = 215,
P = 4.9e-14

TREMBL:DR41081_1 product: "b-catenin"; *Danio rerio* b-catenin mRNA,
complete cds., N = 1, Score = 195, P = 5.8e-12

>PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*)
Length = 578

HSPs:

Score = 237 (35.6 bits), Expect = 7.8e-17, P = 7.8e-17
Identities = 106/401 (26%), Positives = 177/401 (44%)

```

Query: 92 AGGIPLLARLLKTSHEMMLIPVVGTLQECASEENYRAAIKAERIIENLVKNLSENELQ 151
      +GG PL A      +N+ +      L      E Y      + E ++E ++ L S++ Q+Q
Sbjct: 45 SGG-PLKALTTLVYSNQLNLSAALFAEITEKYVRQVSRE-VLEPILILLQSQDPQIQ 102

Query: 152 EHCAMAIYQCAEDKETRD LRLHGGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENV 211
      A+ A + E + L+ GGL+PL + + DN E      G I      + +N
Sbjct: 103 VAACALGNLAVNNENKLLIVEMGGLEPLINQMMG-DNVEVQCNVGCITNLATRDDNKH 161

Query: 212 KFREYKAIE TLVGLLTDQP EEVLNVVGA LGCECCQERENRVIVRKCGGIQPLVNLLVGIN 271
      K      A+ L L      + V N GAL      ENR + G + LV+LL +
Sbjct: 162 KIATSGALIP LTKLAKSKHVRQNRATGALLNMTHSEENRKELVNAGAVPVVLSLSSTD 221

Query: 272 QALLVNVTKAVGACAVEPESMMIIDRLDG--VRLWLSLLKNPHPDVKASAAWALCPCIKN 329
      +      T A+ AV+ + + + + V L SL+ +P VK A AL      +

```

Sbjct: 222 PDVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPPSSRVKQATLALRNLASD 281

Query: 330 AKDAGEMVRSFVGGLELIVNLLKSDNKE-VLASVCAITNIAKDQENLAVITDHGVV-PL 387
E+VR+ GGL +V L++SD+ VLASV A I NI+ N +I D G + PL

Sbjct: 282 TSYQLEIVRA--GGLPHLVKLIQSDSIPLVLASV-ACIRNISIHPLNEGLIVDAGFLKPL 338

Query: 388 LSKLANTNNKLRHHLAEAISRCCMWG-RNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQ 446
+ L ++ +++ H + +NR F E AV + +V ++

Sbjct: 339 VRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSV-QSEIS 397

Query: 447 ALYQLSEAD-NCITMHENGAVKLLLDVMGSPDQDLQEAAGCISNI 492
A + + AD + + + E + L+ M S +Q++ AA ++N+

Sbjct: 398 ACFAILALADVSKLDLLEANILDALIPMTFSQNEVSGNAAAALANL 444

Score = 213 (32.0 bits), Expect = 3.6e-14, P = 3.6e-14
Identities = 81/341 (23%), Positives = 163/341 (47%)

Query: 163 EDKETRDVRLHGGGLKPLASLLNNTD-NKERLAAVTGAIWKCSISKENVTKFREYKAIET 221
EDK+ D G LK L +L+ + + N +R AA+ A I+++ V + + +E

Sbjct: 36 EDKQDLDFYS-GGPLKALTTLVYSDNLLQRSAAALFA----EITEKYVRQVSR-EVLEP 89

Query: 222 LVGLLTQDPEEVLVNVVVGALGECCQERENRVIRKCGGIQPLVNLVGINQALLVNVTKA 281
++ LL Q ++ V ALG EN++++ + GG++PL+N ++G N + N

Sbjct: 90 ILILLQSDPQIQVAACAALGNLAVNNENKLLIVEMGGLEPLINQMMDNVEVQCNAVGC 149

Query: 282 VGACAVEPESMMIIDRLDGVRLLSLLKNPHPDVKASAALWCPCIKNAKDAGEMVRSFV 341
+ A ++ I + L L K+ H V+ +A AL + ++ E+V +

Sbjct: 150 ITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA-- 207

Query: 342 GGLELIVNLLKSDNKEVLASVCAITNIAKDQENLAVI--TDHGVVPLLSKLANTNNNKL 399
G + ++V+LL S + +V A++NIA D+ N + T+ +V L L ++ +++

Sbjct: 208 GAVPVLVSLSSSTDPOVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPPSSRV 267

Query: 400 RHHLAEAISRCCMWGRNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQALYQLSEADNCI 459
+ A+ ++ + LV+ ++S+ + A+ + +S N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query: 460 TMHENGAVKLLLDVMGSPDQDLQEAAGCISNIRRLALATEKAR 503
+ + G +K L+ ++ D + E +S +R LA ++EK R

Sbjct: 328 LIVDAGFLKPLVRLLDYKDE--EIQCHAVSTLRNLAASSEKNR 369

Score = 180 (27.0 bits), Expect = 1.6e-10, P = 1.6e-10
Identities = 80/346 (23%), Positives = 142/346 (41%)

Query: 145 SENEQLQEHCMAMAIYQCAEDKETRDVRLHGGGLKPLASLLNNTDNKERLAAVTGAIWKCS 204
S+N LQ A+A + E K R + R L+P+ LL + D + ++AA A+ +

Sbjct: 58 SDNLLNQRSAAALFAEITE-KYVRQVSR--EVLEPILILLQSDPQIQVAACA-ALGNLA 113

Query: 205 ISKENVTKFREYKAIETLVGLLTQDPEEVLVNVVVGALGECCQERENRVIRKCGGIQPLV 264
++ EN E +E L+ + EV N VG + +N+ + G + PL

Sbjct: 114 VNNENKLLIVEMGGLEPLINQMMDNVEVQCNAVGCITNLATRDDNKHKIATSGALIPLT 173

Query: 265 NLLVGINQALLVNVTKAVGACAVEPESMMIIDRLDGVRLLSLLKNPHPDVKASAALWC 324
L + + N T A+ E+ + V +L SLL + PDV+ AL

Sbjct: 174 KLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLSSSTDPOVQYYCTTALS 233

Query: 325 PCIKNAKDAGEMVRSFVGGLELIVNLLKSDNKEVLASVCAITNIAKDQENLAVITDHGV 384
+ + ++ ++ + +V+L+ S + V A+ N+A D I G

Sbjct: 234 NIAVDEANRKKLAQTEPRLVSKLVSLMDSPPSSRVKQATLALRNLASDTSYQLEIVRAGG 293

Query: 385 VPLLSKLANTNNKLRHHLAEAISRCCMWGRNRVAFGEHKAVAPLVRYLKSNDTNVHRAT 444
+P L KL +++ L I + N + + PLVR L D+ +

Sbjct: 294 LPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDEEIQCH 353

Query: 445 A-QALYQLSEAD-NCITMHENGAVKLLLDVMGSPDQDLQEAAGCIS 490
A L L+ ++ N E+GAV+ ++ +Q + C +

Sbjct: 354 AVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFA 401

Score = 155 (23.3 bits), Expect = 8.8e-08, P = 8.8e-08
Identities = 88/401 (21%), Positives = 175/401 (43%)

Query: 60 LYEARD--VEVARCGALALWSCSKSHTNKEAIRKAGGI-PLLARLLKTSHENMLIPVVG 116
L +++D ++VA C AL + + ++ NK I + GG+ PL+ ++ + E + VG

Sbjct: 93 LLQSDPQIQVAACAALG--NLAVNNENKLLIVEMGGLEPLINQMMDNVE-VQCNVGC 149

Query: 117 LOECASEENYRAAIKAERIIENLVKNLNSENEQLQEHCMAMAIYQCAEDKETR-DLVRLHG 175
+ A+ ++ + I + L K S++ ++Q + A+ +E R +LV G

Sbjct: 150 ITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA-G 208

Query: 176 GLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFR--EYKAIETLVGLLTQDPEEV 233
+ L SLL++TD + T A+ ++ + N K E + + LV L+ V

Sbjct: 209 AVPVLVSLSSSTDPDVQYYCTT-ALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPPSSRV 267

Query: 234 LVNVVVGALGECCQERENRVIVRCKGGIQLVNLVGINQALLVNVTKAVGACAVEPESMM 293
AL + ++ + GG+ LV L+ + L++ + ++ P +

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSOSIPLVLASVACIRNISIHPLNEG 327

Query: 294 IIDRLDGVRLWLSLK-NPHPDVKASAAWALCPCIKNA-KDAGEMVRSFVGGLELIVNLL 351
+I ++ L LL +++ A L ++ K+ E S G+E L

Sbjct: 328 LIVDAGFLKPLVRLLDYKDSEEIQCCHAVSTLRNLAASSEKNRKEFFES--GAVEKCKELA 385

Query: 352 KSDNKEVLA--SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNNKLRRHHLAEAISR 409
V + S C AI +A D L ++ + ++ L + + N ++ + A A++

Sbjct: 386 LDSPVSVQSEISACFAILALA-DVSKDLL-EANILDALIPMTFSQNEVSGNAAAALAN 443

Query: 410 CCMWGRNRVAFGE-----HKAVAP-LVRYLKSNDTVHRATAQALYQLSE 453
C N E ++ + L+R+LKS+ + QL E

Sbjct: 444 LCSRVNNYTKIIEAWDRPNEGIRGFLIRFLKSDYATFEHIALWTILQLE 493

Score = 139 (20.9 bits), Expect = 5.0e-06, P = 5.0e-06
Identities = 80/329 (24%), Positives = 142/329 (43%)

Query: 37 GGITKLVALLDCAHD-STKPAQ---SSLYEARDVEVARCGALALWSCSKSHTNKEAIRKA 92
G I T L D H +T A + L +++ + V R AL + + S N++ + A

Sbjct: 148 GCITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA 207

Query: 93 GGIPLLARLLKTSHEENMLIPVVGTLQECASEE-NYRAAIKAE-RIENLVKNLNSENEQL 150
G +P+L LL ++ ++ L A +E N + + E R++ LV ++S + ++

Sbjct: 208 GAVPVLVSLSSSTDPDVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPPSSRV 267

Query: 151 QEHCAMAIYQCAEDKETR-DLVRHLHGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKEN 209
+ +A+ A D + ++VR GGL L L+ + D+ + A I SI N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRA-GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLN 325

Query: 210 VTKFREYKAIETLVGLLT-DQPEEVLVNVVVGALGECCQERE-NRVIVRCKGGIQLVNL 267
+ ++ LV LL EE+ + V L E NR + G ++ L

Sbjct: 326 EGLIVDAGFLKPLVRLLDYKDSEEIQCCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELA 385

Query: 268 VG--INQALLVNVTKAVGACA-VEPESMMIIDRLDGVRLWLSLLKNPHPDVKASAAWA-L 323
+ ++ ++ A+ A A V ++ + LD + + + +N A+AA A L

Sbjct: 386 LDSPVSVQSEISACFAILALADVSKDLLLEANILDAL-IPMTFSQNEVSGNAAAALANL 444

Query: 324 CPCIKN-AKDAGEMVRSFVGGLELIVNLLKSD 354
C + N K R G ++ LKSD

Sbjct: 445 CSRVNNYTKIIEAWDRPNEGIRGFLIRFLKSD 476

Score = 136 (20.4 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 72/304 (23%), Positives = 133/304 (43%)

Query: 58 SSLYEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEENMLIPVVGTL 117
+ L +++ + V R AL + + S N++ + AG +P+L LL ++ ++ L

Sbjct: 173 TKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLSSSTDPDVQYYCTTAL 232

Query: 118 QECASEE-NYRAAIKAE-RIENLVKNLNSENEQLQEHCAMAIYQCAEDKETR-DLVRHL 174
A +E N + + E R++ LV ++S + ++ +A+ A D + ++VR

Sbjct: 233 SNIADVDEANRKKLAQTEPRLVSKLVSLMDSPPSSRVKCQATLALRNLASDTSYQLEIVRA- 291

Query: 175 GGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLT-DQPEEV 233
GGL L L+ + D+ + A I SI N + ++ LV LL EE+

Sbjct: 292 GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDSEI 350

Query: 234 LVNVVVGALGECCQERE-NRVIVRCKGGIQLVNLVVG--INQALLVNVTKAVGACA-VEP 289
+ V L E NR + G ++ L + ++ ++ A+ A A V

Sbjct: 351 QCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFAILALADVSK 410

Query: 290 ESMIIIDRLDGVRLWLSLLKNPHPDVKASAAWA-LCPCIKN-AKDAGEMVRSFVGGLELI 347
++ + LD + + + +N A+AA A LC + N K R G +

Sbjct: 411 LDLEANILDAL-IPMTFSQNEVSGNAAAALANLCSRVNNYTKIIEAWDRPNEGIRGFL 469

Query: 348 VNLLKSD 354
+ LKSD

Sbjct: 470 IRFLKSD 476

Score = 114 (17.1 bits), Expect = 2.7e-03, P = 2.7e-03
Identities = 71/335 (21%), Positives = 132/335 (39%)

Query: 1 MVNILDSPHKSCLKLAAETIANVAKFKRRARRVVRQHGGITKLVALLDCAHDSTKPAQSSL 60
+ + S H ++ A + N+ + R+ + G + LV+LL ST P

Sbjct: 172 LTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLSS-----STDP----- 222

Query: 61 YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEENMLIPVVGTLQEC 120
DV+ AL+ + +++ K A + + L L+ + + L+

Sbjct: 223 ----DVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPPSRVKCQATLALRNL 278

Query: 121 ASEENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETRDVLRLHGGLKPL 180
AS+ +Y+ I + +LVK + S++ L I + L+ G LKPL

Sbjct: 279 ASOTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPL 338

Query: 181 ASLLNNTDNKERLAAVTGAIWKCSISKE-NVTKFREYKAIETLVGLLTDQPEEVLNVVVG 239
LL+ D++E + + S E N + F E A+E L D P V +

Sbjct: 339 VRLLDYKDSEIIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISA 398

Query: 240 ALGECCQERENRVIRKCGGIQPLVNLVLGINQALLVNVTKAVG-ACAVEPESMMIIDRL 298
+++ + + + L+ + NQ + N A+ C+ II+

Sbjct: 399 CFAILALADVSKLDLLEANILDALIPMTFSQNEQVSGNAAAALNLCRSRVNNYTKIIEAW 458

Query: 299 D----GVR-LLWSLLKNPHPDVKASAAWALCPCIKNAKDAGE 335
D G+R L LK+ + + A W + +++ D E

Sbjct: 459 DRPNEGIRGFLIRFLKSDYATFEHIALWTILQLLESHNDKVE 500

Score = 106 (15.9 bits), Expect = 2.0e-02, P = 2.0e-02
Identities = 49/204 (24%), Positives = 89/204 (43%)

Query: 65 DVEVARCGALA-LWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEMLIPVVGTLQECA-S 122
+VEV +C A+ + + + NK I +G + L +L K+ H + G L S

Sbjct: 139 NVEV-QCNAVGCITNLATRDDNKKKIATSGALIPLTAKSKHIRVQRNATGALLNMTHS 197

Query: 123 EENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETRD-LVRLHGGL-KPL 180
EEN + + A + LV L+S + +Q +C A+ A D+ R L + L L

Sbjct: 198 EENRKELVNAGAV-PVLVSLLSSTDPDVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKL 256

Query: 181 ASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLNVVVG 240
SL+++ ++ + A T A + + + LV L+ +++ V

Sbjct: 257 VSLMDSPPSRVKCQA-TLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVAC 315

Query: 241 LGECCQERENRVIRKCGGIQPLVNL 267
+ N ++ G ++PLV LL

Sbjct: 316 IRNISIHLPLNEGLIVDAGFLKPLVRL 342

Pedant information for DKFZphtes3_35p17, frame 3

Report for DKFZphtes3_35p17.3

[LENGTH] 505
[MW] 55224.34
[pI] 8.43
[HOMOL] PIR:S50446 VAC8 protein - yeast (Saccharomyces cerevisiae) 2e-16
[FUNCAT] 30.25 vacuolar and lysosomal organization [S. cerevisiae, YEL013w] 8e-18
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w] 8e-18
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YEL013w] 8e-18
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YNL189w] 3e-06
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL189w] 3e-06
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL189w] 3e-06
[BLOCKS] BL01265C
[BLOCKS] BL00242A Integrins alpha chain proteins
[SCOP] d3bct_1.91.1.1.1 beta-Catenin [Mouse (Mus musculus) 7e-18
[PIRKW] cytosol 3e-11
[PIRKW] apoptosis 3e-11
[PIRKW] carcinogenesis 3e-11
[PIRKW] cell adhesion 3e-11
[PIRKW] cytoskeleton 3e-12
[SUPFAM] pendulin 1e-07
[KW] All_Alpha
[KW] 3D
[KW] LOW_COMPLEXITY 2.38 %

SEQ MVNILDSPHSLKCLAAETIANVAKFKRARRVVRQHGGITKLVALLDCAHDSTKPAQSSL
SEGXXXXXXXXXXXXX.....
2bct-HH

SEQ YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEMLIPVVGTLQEC
SEG
2bct- HHCCCHHHHHHHHHHHHHHHCHHHHHHHHHCHHHHHHHHHGGGCCCHHHHHHHHHHHHHHH

SEQ ASEENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETRDVLRLHGGLKPL
SEG
2bct- HHTTTTHHHHHHHHHCHHHHHHHHHCHHHHHHHHHHHHHHHHTTHHHHHHHHHCHHHHHHH

```
SEQ  ASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLVNVVGA
SEG  .....
2bct- HHHHH-HCCCHHHHHHHHHHHHHHCCCHHHHHHHHHCHHHHHHTTTTCCCHHHHHHHHHH

SEQ  LGECQCQERENRVIVRKCGGIQPLVNLVGINQALLVNVTKAVGACAVEPESMMIIDRLDG
SEG  .....
2bct- H-----HHHHHCCCCCTTTHHHHHHHHHHHHCTTTHHHHHHHHTTTTHHHHHHH-HHCH

SEQ  VRLWSLLKNPHPDVKASAAWALCPCIKNAKDAGEMVRSFVGLELIVNLLKSDNKEVLA
SEG  .....
2bct- HHHHHHHHTTTTHHHHHHHHHHHHHHCCCHH-HHHHHHHHHHHHHHCTTTTTHHHH

SEQ  SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEAISRCMWGRNRVAF
SEG  .....
2bct- HHHHHHHHHHCGGGHHHHHHCHHHHHHHHHHHHTTCCCHHHHHHHHHCHHHHH

SEQ  GEHKAVAPLVRYLKSNDTNVHRATAQALYQLSEADNCITMHENGAVKLLLDVGS PDQD
SEG  .....
2bct- HTTTTHHHHHHHCCCCCHHHHHHHHHHHHTTTHHHHHHHHCCCHHHHHHTTTTTHH

SEQ  LQEAAAGCISNIRRLALATEKARYT
SEG  .....
2bct- HHHHHHHH.....
```

(No Prosite data available for DKFZphtes3_35p17.3)

(No Pfam data available for DKFZphtes3_35p17.3)

DKFZphtes3_35p22

group: cell cycle

DKFZphtes3_35p22 encodes a novel 549 amino acid protein, with similarity to oncogene 1 (tre-2 locus).

The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control.

The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

strong similarity to oncogene 1 (tre-2 locus)

membrane regions: 1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: map="17"

Insert length: 2072 bp

Poly A stretch at pos. 2062, polyadenylation signal at pos. 2039

```
1 GTTACACACA GGCAGTGGTA TCTGTGAGCA GCTCTGTGGA CTCAAAGGTT
51 TTCTCCCTGA GAGGCATGAC CCAGGCCAGC TGATTCATCA GAATCAGGAT
101 GGACGTGGTA GAGGTCGCGG GCAGTTGGTG GGCACAAGAG CGAGAGGACA
151 TCATTATGAA ATACGAAAAG GGACACCGAG CTGGGCTGCC AGAGGACAAG
201 GGGCCTAAGC CTTTTCGAAG CTACAACAAC AACGTCGATC ATTTGGGGAT
251 TGTACATGAG ACGGAGCTGC CTCCTCTGAC TGCGCGGGAG GCGAAGCAAA
301 TTCGGCGGGA GATCAGCCGA AAGAGCAAGT GGGTGGATAT GCTGGGAGAC
351 TGGGAGAAAT ACAAAGCAG CAGAAAGCTC ATAGATCGAG CGTACAAGGG
401 AATGCCCCATG AACATCCGGG GCCCGATGTG GTCAGTCCTC CTGAACACTG
451 AGGAAATGAA GTTGAAAAC CCCGGAAGAT ACCAGATCAT GAAGGAGAAG
501 GGCAAGAAGT CATCTGAGCA CATCCAGCGC ATCGACCGGG ACCTAAGCGG
551 GACATTAAGG AAGCATATAT TCTTCAGGGA TCGATACGGA ACCAAGCAGC
601 GGGAACTACT CCACATCCTC CTGGCATATG AGGAGTACAA CCCGGAGGTG
651 GGCTACTGCA GGGACCTGAG CCACATCGCC GCCTTGTTCC TCCTCTATCT
701 TCCTGAGGAG GATGCATTCT GGGCACTGGT GCAGCTGCTG GCCAGTGAGA
751 GGGCACTCCCT GCAGGGGATT CACAGCCCAA ATGGCGGGAC CGTCCAGGGG
801 CTCGAAGACC AACAGGAGCA TGTGGTAGCC ACGTCACAAC CCAAGACCAT
851 GGGGCATCAG GACAAGAAAG ATCTATGTGG GCAGTGTGCC CCGTTAGGCT
901 GCCTCATCCG GATATTGATT GACGGGATCT CTCTCGGGCT CACCCTGCGC
951 CTGTGGGAGC TGTATCTGGT AGAAGGCGAA CAGGCGCTGA TGCCGATAAC
1001 AAGAATCGCC TTTAAGGTTT AGCAGAAGCG CCTCACGAAG ACGTCCAGGT
1051 GTGGCCCGTG GGCACGTTTT TGCAACCGGT TCGTTGATAC CTGGGCCAGG
1101 GATGAGGACA CTGTGCTCAA GCATCTTAGG GCCTCTATGA AGAACTAAC
1151 AAGAAAGAGG GGGGACCTGC CACCCCCAGC CAAACCCGAG CAAGGGTCGT
1201 CGGCATCCAG GCCTGTGCGG GCTTCACGTG GCGGGAAGAC CCTCTGCAAG
1251 GGGGACAGGC AGGCCCTCC AGGCCACCA GCCCGGTTCC CGCGGCCCAT
1301 TTGGTCAGCT TCCCGCCAC GGGCACCTCG TTCTTCCACA CCCTGTCTTG
1351 GTGGGCTGTG CCGGGAAGAC ACCTACCCTG TGGGCACTCA GGGTGTGCCC
1401 AGCCCCGGCC TGGCTCAGGG AGGACCTCAG GGTTCCTGGA GATTCTTGCA
1451 GTGGAATCC ATGCCCGGCC TCCCAACGGA CCTGGACGTA GAGGGCCCTT
1501 GGTTCGCGCA TTATGATTTC AGACAGAGCT GCTGGGTCCG TGCCATATCC
1551 CAGGAGGACC AGCTGGCCCC CTGCTGGCAG GCTGAACACC CTGCGGAGCG
1601 GGTGAGATCG GCTTTCGCTG CACCCAGCAC TGATTCCGAC CAGGGCACCC
1651 CTTTCAGAGC TAGGGACGAA CAGCAGTGTG CTCCCACCTC AGGGCCTTGC
1701 CTCTGCGGCC TCCACTTGGA AAGTCTCAG TTCCCTCCAG GCTTCTAGAA
1751 GCATCTGGGC CAGGGCTCAT GGCTGGATAA TTTCCCTAGG CTTAACAACC
1801 CAAGCAAGCT TCGCATCCTC GTTTTATTTT TGGTTAAACT TATGAAAATG
1851 TATTAAGAAA GAGTGCAGCT CGAGAGAGAT TCAGAGATGG AACACACCAG
1901 ACCCCAGATC ACAAAGCCAA CCATGCCAG CCCCTCCAG CACCCACAGC
1951 CCCACGACCA TCGTTCTGAA TTCTGACGAC ACCGTGAGCC TGCCTTTGTA
2001 CTTCAAATCT ATGGAAGGAT AACCACCTTC ATGTTTGA AATAATGTTT
2051 CCTGTTGAAA TGAAAAAAA AA
```

BLAST Results

Entry AC003976 from database EMBL:
Homo sapiens chromosome 17, clone hCIT.91_J_4, complete sequence.
Score = 4385, P = 0.0e+00, identities = 881/886

14 exons

Entry HSG19723 from database EMBL:
human STS A001W35.

Score = 850, P = 1.9e-32, identities = 170/170

Medline entries

92228503:

A novel transcriptional unit of the tre oncogene widely expressed in human cancer cells.

94067315:

The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene.

95176708:

UBP5 encodes a putative yeast ubiquitin-specific protease that is related to the human Tre-2 oncogene product.

Peptide information for frame 3

ORF from 99 bp to 1745 bp: peptide length: 549
Category: strong similarity to known protein

```

1 MDVVEVAGSW WAQEREDIIM KYEKGHRAGL PEDKGPKPFR SYNNNVVDHLG
51 IVHETELPPL TAREAKQIRR EISRKSKWVD MLGDWEKYKS SRKLIDRAYK
101 GMPMNIRGPM WSVLLNTEEM KLKNPGRYQI MKEKGKKSSE HIQRIDRDVS
151 GTLRKHIFFR DRYGTKQREL LHILLAYEY NPEVGYCRDL SHIAALFLLY
201 LPEEDAFWAL VQLLASERHS LQGFHSPNGG TVQGLQDQQE HVVATSQPKT
251 MGHQDKKDLG GQCSPLGCLI RILIDGISLG LTLRLWDVYL VEGEQALMPI
301 TRIAFKVQOK RLTKTSRCGP WARFCNREVD TWARDLTVL KHLRASMKKL
351 TRKKGDLPPP AKPEQGSAS RPVPASRGGK TLCKGDRQAP PGPPARFPRP
401 IWSASPPRAP RSSTPCPGA VREDTYPVGT QGVPSPALAQ GGPOGSRFL
451 QWNSMPRLPT DLDVEGPWFR HYDFRQSCWV RAISQEDQLA PCWQAEHPAE
501 RVRSAFAAPS TDSQGTFFR ARDEQQCAPT SGPCLCGLHL ESSQFPFGF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_35p22, frame 3

PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human, N = 1, Score = 2181, P = 5.5e-226

PIR:S57867 oncogene 1 - human, N = 1, Score = 1536, P = 1.2e-157

>PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human
Length = 786

HSPs:

Score = 2181 (327.2 bits), Expect = 5.5e-226, P = 5.5e-226
Identities = 405/500 (81%), Positives = 440/500 (88%)

```

Query: 1 MDVVEVAGSWWAQEREDIIMKYEKGHRAGLPEDKGPKPFRSYNNNVVDHLGIVHETELPPL 60
MD+VE A S AQER+DI+MKY+KGRAGLPEDKGP+P N+++D GI+HETELPP+
Sbjct: 1 MDMVENADSLQAQERKDILMKYDKGHRAGLPEDKGPEPV-GINSSIDRFILHETELPPV 59

Query: 61 TAREAKQIRREISRKSKWVDMLGDWEKYKSSRKLIIDRAYKGMMPMNIRGPMWSVLLNTEEM 120
TAREAK+IRRE++R SKW++MLG+WE YK S KLIDR YKG+PMNIRGP+WSVLLN +E+
Sbjct: 60 TAREAKKIRREMTRTSKWMEMLGEWETKYHSSKLIDRVYKGI+PMNIRGPVWSVLLNIQEI 119

Query: 121 KLKNPGRYQIMKEKGKKSSEHIQRIDRDVSGTLRKHIFFRDRYGTKQRELLHILLAYEY 180
KLKNPGRYQIMKE+GK+SSEHI ID DV TLR H+FFRDRYG KQREL +ILLAY EY
Sbjct: 120 KLKNPGRYQIMKERGRSSEHIHIDLDVRTTLRNHVFFRDRYGAQRELFYILLAYSEY 179

Query: 181 NPEVGYCRDLSHIAALFLLYLPEEDAFWALVQLLASERHSLQGFHSPNGGTVQGLQDQQE 240
NPEVGYCRDLSHI ALFLLYLPEEDAFWALVQLLASERHSL GFHSPNGGTVQGLQDQQE
Sbjct: 180 NPEVGYCRDLSHITALFLLYLPEEDAFWALVQLLASERHSLPGFHSPNGGTVQGLQDQQE 239

```

Query: 241 HVVATSQPKTMGHQDKKDLGGCQCSPLGCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI 300
HVV SQPKTM HQDK+ LCGQC+ LGCL+R LIDGISLGLTLRLWDVYLVEGEQ LMPI
Sbjct: 240 HVVPKSSQPKTMWHQDKEGLCGQCASLGCLLRNLIDGISLGLTLRLWDVYLVEGEQVLMPI 299

Query: 301 TRIAFKVVQKRLTKTSRCGPWARFCNRFVDTWARDEDTVLKHLRASMKKLTRKKGDLPPP 360
T IA KVQKRL KTSRCG WAR N+F DTWA ++DTVLKHLRAS KKLTRK+GDLPPP
Sbjct: 300 TSIALKVQKRLMKTSRCGLWARLRNQFFDTWAMNDDTVLKHLRASTKKLTRKQGDLP 359

Query: 361 AKPEQGSSASRPVPASRGGKTLCKGDRQAPPGPAPRFPRIWSASPPRAPSSTPCPGGA 420
AK EQGS A RPVPASRGGKTLCKG RQAPPGPPA+F RPI SASPP A R STPCPGGA
Sbjct: 360 AKREQGSLAPRPVPASRGGKTLCKGYRQAPPGPPAQFQRPICASPPWASRFSTPCPGGA 419

Query: 421 VREDTYPVGTQGVPSPALAQGGPQGSWRFLQNSMPRLPTDLDVEGPWFRHYDFRQSCWV 480
VREDTYPVGTQGVPS ALAQGGPQGSWRFL+W SMPRLPTDLD+ GPWF HYDF +SCWV
Sbjct: 420 VREDTYPVGTQGVPSLALAQGGPQGSWRFLQNSMPRLPTDLDIGGPWFPHYDFERSCWV 479

Query: 481 RAISQEDQLAPCWAQEAHPAE 500
RAISQEDQLA CWAQEAH E
Sbjct: 480 RAISQEDQLATCWAQEAHCGE 499

Pedant information for DKFZphtes3_35p22, frame 3

Report for DKFZphtes3_35p22.3

[LENGTH] 549
[MW] 62159.16
[pI] 9.23
[HOMOL] PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human 0.0
[FUNCAT] 11.01 stress response [S. cerevisiae, YGR100w] 2e-16
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YGR100w] 2e-16
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YNL293w] 3e-15
[PIRKW] transmembrane protein 6e-14
[PROSITE] MYRISTYL 6
[PROSITE] AMIDATION 1
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 10
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 5.28 %

SEQ MDVVEVAGSWSWAQEREDIIMKYERGHRAGLPEDKGPKPFRSYNNVNDHLGIVHETELPPL
SEG
PRD cccceccchhhhhhhhhhhhhhhcc
MEM

SEQ TAREAKQIRREISRSKSWVDMLGDWEKYSSRKLIDRAYKGMPMNIRGPMWSVLLNTEEM
SEG
PRD chhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhcccccccccccccccccccc
MEM

SEQ KLKNPGRYQIMKEKGKKSSEHIQRIDRDVSGTLRKHIFFRDYGTQRELHILLAYEY
SEG
PRD cccccchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc
MEM

SEQ NPEVGYCRDLSHIAALFLLYLPEEDAFWALVQLLASERHSLQGFHSPNGGTVQGLQDQOE
SEG
PRD cccccccccchhh
MEM

SEQ HVVATSQPKTMGHQDKKDLGGCQCSPLGCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI
SEG
PRD hhhhhhhchhhhhhhhhccccccccchhhhhhhhhhhccccchhhhhhhhhhhccccceeehh
MEMMMMMMMMMMMMMMMMM.....

SEQ TRIAFKVVQKRLTKTSRCGPWARFCNRFVDTWARDEDTVLKHLRASMKKLTRKKGDLPPP
SEG
PRD hhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM

SEQ AKPEQGSSASRPVPASRGGKTLCKGDRQAPPGPAPRFPRIWSASPPRAPSSTPCPGGA
SEG
PRD ccc
MEM


```

SEQ  VREDTYPVGTQGVPSPALAQGGPQGSWRFLQWNSMPRLPTDLDVEGPWFRHYDFRQSCWV
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  RAISQEDQLAPCWQAEHPAERVRSFAFAAPSTDSDQGTFFRARDEQQCAPTSGPCLCGLHL
SEG  .....
PRD  cchhhhhhhhhhhhhcchhhhhhhhhccccccccccccccccchhhhhccccccccccceeee
MEM  .....

SEQ  ESSQFP PGF
SEG  .....
PRD  ccccccccc
MEM  .....

```

Prosites for DKFZphtes3_35p22.3

PS00004	136->140	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	310->314	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	348->352	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	73->76	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	152->155	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	446->449	PKC_PHOSPHO_SITE	PDOC00005
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00007	93->100	TYR_PHOSPHO_SITE	PDOC00007
PS00007	92->100	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	101->107	MYRISTYL	PDOC00008
PS00008	230->236	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	441->447	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_35p22.3)

DKFZphtes3_4b4

group: testes derived

DKFZphtes3_4b4 encodes a novel 497 amino acid protein similar to SCP proteins and a human trypsin inhibitor.

The novel protein contains an extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2, predicted by Prosite and Pfam. This domain is found in a variety of extracellular proteins from eukaryotes that have been found to be evolutionary related. The exact function of these proteins is not yet known. In addition, the protein is similar to a human trypsin inhibitor.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes or as a new protease inhibitor.

strong similarity to trypsin inhibitor

might be a new protease inhibitor?

Sequenced by AGOWA

Locus: /map="333.4 cR from top of Chr16 linkage group"

Insert length: 4574 bp

Poly A stretch at pos. 4551, polyadenylation signal at pos. 4539

```

1  GGGCGGTGCT CCCATTGAGC TGTCTGCTCG CTGTGCCCCG TGTGCCTGCT
51  GTGCCCGCGC TGTGCGCCGT GCTACCGCGT CTGCTGGACG CGGGAGACGC
101  CAGCGAGCTG GTGATTGGAG CCCTGCGGAG AGCTCAAGCG CCCAGCTCTG
151  CCCGAGGAGC CCAGGCTGCC CCGTGAGTCC CATAGTTGCT GCAGGAGTGG
201  AGCCATGAGC TGCCTCCTGG GTGGTGTCAT CCCCTTGGGG CTGCTGTTCC
251  TGGTCTGCGG ATCCCAAGGC TACCTCCTGC CCAACGTCAC TCTCTTAGAG
301  GAGCTGCTCA GCAAAATACCA GCACAACGAG TCTCACTCCC GGGTCCGCAG
351  AGCCATCCCC AGGGAGGACA AGGAGGAGAT CCTCATGCTG CACAACAAGC
401  TTCGGGGGCA GGTGCAGCCT CAGGCCCTCA ACATGGAGTA CATGACCTGG
451  GATGACGAAC TGGAGAAGTC TGCTGCAGCG TGGGCCAGTC AGTGCATCTG
501  GGAGCACGGG CCCACCACTG TGCTGGTGTC CATCGGGCAG AACCTGGGCG
551  CTCACCTGGG CAGGTATCGC TCTCCGGGGT TCCATGTGCA GTCTCTGGTAT
601  GACGAGGTGA AGGACTACAC CTACCCCTAC CCGAGCGAGT GCAACCCCTG
651  GTGTCCAGAG AGGTGCTCGG GGCCTATGTG CACGCACTAC ACACAGATAG
701  TTTGGGCCAC CACCAACAAG ATCGGTTGTG CTGTGAACAC CTGCCGGAAG
751  ATGACTGTCT GGGGAGAAAG TTGGGAGAAC GCGGTCTACT TTGTCTGCAA
801  TTATTCTCCA AAGGGGAAGT GGATTGGAGA AGCCCCCTAC AAGAATGGCC
851  GGCCCTGCTC TGAGTGCCCA CCCAGCTATG GAGGCAGCTG CAGGAACAAC
901  TTGTGTTACC GAGAAGAAAC CTACACTCCA AAACCTGAAA CGGACGAGAT
951  GAATGAGGTG GAAACGGCTC CCATTCTCTG AGAAAACCAT GTTTGGCTCC
1001  AACCGAGGGT GATGAGACCC ACCAAGCCCA AGAAAACCTC TCGGGTCAAC
1051  TACATGACCC AAGTCGTGAG ATGTGACACC AAGATGAAGG ACAGGTGCAA
1101  AGGGTCCACG TGTAACAGGT ACCAGTGCCC AGCAGGCTGC CTGAACCACA
1151  AGCGCAAGAT CTTTGGAAGT CTGTTCTATG AAAGCTCGTC TAGCATATGC
1201  CGCGCCGCCA TCCACTACGG GATCCTGGAT GACAAGGGAG GCCTGGTGGA
1251  TATCACCAGG AACGGGAAGG TCCCTTCTT CGTGAAGTCT GAGAGACACG
1301  GCGTGCAATC CCTCAGCAAA TACAAACCTT CCAGCTCATT CATGGTGTCA
1351  AAAGTGAAAG TGCAGGATTT GGACTGCTAC ACGACCGTTG CTCAGCTGTG
1401  CCCGTTTGAA AAGCCAGCAA CTCACTGCCC AAGAATCCAT TGTCCGGCAC
1451  ACTGCAAGA GCAACCTTCC TACTGGGCTC CGGTGTTTGG AACCAACATC
1501  TATGCAGATA CCTCAAGCAT CTGCAAGACA GCCGTGCACG CGGGAGTCAT
1551  CAGCAACGAG AGTGGGGGTG ACGTGGACGT GATGCCCGTG GATAAAAAGA
1601  AGACCTACGT GGGCTCGCTC AGGAATGGAG TTCAGTCTGA AAGCCTGGGG
1651  ACTCCTCGGG ATGGAAGGCG CTTCCGGATC TTTGCTGTCA GGCAGTGAAT
1701  TTCCAGCACC AGGGGAGAAG GGGCGTCTTC AGGAGGGCTT CGGGGTTTGG
1751  CTTTTATTTT TATTTTGTC TATGCGGGTA TATGGAGAGT CAGGAACTT
1801  CCTTTGACTG ATGTTTCAAG TCCATCACTT TGTGGCCTGT GGGTGAGGTG
1851  ACATCTCATC CCCTCACTGA AGCAACAGCA TCCCAAGGTG CTCAGCCGGA
1901  CTCCTCTGGT CCTGATCCTG CTGGGGCCCG GGGGTCTCCA TCTGGACGTC
1951  CTCTCTCCTT TAGAGATCTG AGCTGTCTCT TAAAGGGGAC AGTTGCCCAA
2001  AATGTTTCCT GCTATGTGTT CTCTGTGTGG TGGAGGAAGT TGATTTC AAC
2051  CTCCTTGCCA AAAGAACAAA CCATTGGAAG CTCACAATTG TGAAGCATTC
2101  ACGGCGTCGG AAGAGGCCTT TTGAGCAAGC GCCAATGAGT TTCAGGAATG
2151  AAGTAGAAGG TAGTTATTTA AAAATAAAAA ACACAGTCCG TCCCTACCAA
2201  TAGAGGAAAA TGGTTTAAAT GTTTGCTGGT CAGACAGACA AATGGGCTAG
2251  AGTAAGAGGG CTGCGGGTAT GAGAGACCCC GGCTCCGCCC TGGCAGTGTG
2301  CCTTGTCTGG GGGCCGCCAC AGGCCCTTCT CAATGGCCGC ATTGAGGATG
2351  GCTCTATACA CAGCAGTGCT GGTATATGTA GAGTTCAGCA GTCACCTCAG
2401  AGATGTATCT TGTCTTTGTC AGGCCCTTCA TCTTCATGGC CCACCTGTTT
2451  TCTGCCGTGA CCTTTGGTCC CATTGAGGAC TAAGGATCGG GACCCCTTCT

```

```

2501 TTACCCCTTA CCCATTGTGG CTCCCACCCT GCCTCGGACT GGTTTACGTG
2551 TCCTGGTTCA CACCCAGGAC TTTTCTTTGC AAGCGAACCT GTTTGAAGCC
2601 CAAGTCTTAA CTCTGGTCT CGTAAGGTTT CACTGAGACG AGATGTCTGA
2651 GAACAACCAA AGAAGGCCTG CTCTTTGCTG CTTTAAAAA ATGACAATTA
2701 AATGTGCAGA TTCCCCACGC ACCCGATGAC CTATTTTTC AGCCGTGGGA
2751 GGAATGGAGT CTTTGGTACA TTCTCACCG AGGTAGCAG CTCAGTTTGT
2801 GGTATGAAA CCGTCTGTGG CCTCATGACA GCGAGAGATG GGAATACACT
2851 AGAAGGATCT CTTTCTCTGT TTTCTGTAAG CGACTCTTGC CAAACGTTCC
2901 CGAGGCCGCA AGGAGTGTAG TACACCCTGG CTGCCATCAC TCTATAAAAG
2951 TGCTTCATGA GCCCAGACCA AAAGCCACCA GTGAAATGAA GTACCCTTTT
3001 GTAAATAGCA TTTTCTTTC GAAGGTGAAA ATTCACACTCT CTACCACCGG
3051 GCCAGCCAAT AGATCACTTT GGTGAATGCT AGTTTCAAAT TTGATTCAAA
3101 ATATTTCTTA GGTGAAAGAA CTAGCAGAAA GTCAAAACT AAGATACTGT
3151 AGACTGGACA AGAAATCTTA CCTGGGCACC TAGGTGATGC CTTCTTTCTT
3201 TGATTGCCTT TCTAATAAAT GCAGAATCTG AAGGTAAATA GGTTTAAAAAC
3251 AAAACAAAAA CCCACCCCTT TAAGGAGTTG GTAAAAAGCA GTTCAACTCT
3301 TAGCTTGACT GAGCTAAAA TCACAGGACT ACGTGCTTTG TGCATTGTAG
3351 TCTAGTCGTA ATTCATAGGT ACTGACTCCT CAGCCCCAAA TGTGCGGAGAG
3401 GAAGAATTCG GTCAGCCTGT CAGGTCGTGA GTCCAGTTAC CACCAACAT
3451 CTGGGAAACT TCTGGGTGCT GGGTGCTCTG CTGCTGGACT TTTGTGGCTG
3501 TGTCTGTGTC TGCAAGATAA ATTAGATCGC CCTGTGGGGT TTGAGAATT
3551 AGTGAAGGGT CCAGGACGAT CCCAGTGGGC TCGCTTCCAA AGCATCCAC
3601 TCAAGGGAGA CTTGAAACTT CCAGTGTGAG TTGACCCCAT CATTTAAAAA
3651 TAAAGTCCCC GGGTTCCTTA ATGCTCTCTT CACTGGGCCT TCCTAGCAGG
3701 ATAGAAAGTC CTTGCCCGAGA GCAGGACCTG GCTGTCTTTT TTTTCTTTT
3751 TTTCCCGAGA CCAAGTTTCA CTCTGTTGCC CAAGGTAGAG TGCAGTGGCG
3801 TGATCTCTGC TCATTGCAAC TGCCGCTCTC CGGGTTCAG CAATTCTCAT
3851 GCATCAGCCT CCCAAGTACC TGGGACTACA GGCCTGAGCT ACCATGCCCG
3901 GCTAATTTTT GTATTTTATG TAGAGATGGG GTTTCATTAT GTTGGCCAGG
3951 CTGGTCTCGA ACTCCTTACC TCAGGTGATC CACCACCTT GGCCTCCCGA
4001 AGTGCTGGGA TTACAGGCAT GAGCCACTGC GCCCGGCCAT GGACCTGGCT
4051 GTCCTTATCA TCCCCACAAA CATTTTGAAA CTGGAATATT TGTCTTCAGA
4101 AATGGAAC AAGACTATAA ATGATAAGCC CTGTCCCTAG CACCACCTCT
4151 CCTGTGTGTG GAATAGAGGC CCCTCGTGCT ACCAACACTT ACCCTGTGTT
4201 TAAAAAGATC TTGTACCAAG CCAACGGCGT TCCTGGCTCT CCTGCCACA
4251 GGATGAACAT TTTCCGGCTC CTTAGGAGTT TTGCCCTACC GTATTCCAAA
4301 GCGTGTGCTG GTTTCTCATA TTGTCTGTAG GCTCACTCAG CCGCAGTTT
4351 ATGTGTGTGC TTTTCTTAT GAAAAATGAT GTATTTTGCT ACTTCTGTG
4401 TACAAAGTTT TATTGTAAAT GTTTTTTGTG CTTTGCATGA ACAGGGGCCA
4451 CGTTGTGTGA ATTGTTTTCAG TAGAACTGGT TTGATTCTA AAATGTTCTT
4501 GTAACATATC TTTTATGAAC AAATCTGAAC AATTTGTGAA ATAAACATT
4551 GAAACCAAAA AAAAAA AAAA

```

BLAST Results

Entry HS834352 from database EMBL:
human STS WI-15502.

Score = 1331, P = 5.4e-54, identities = 287/301

Medline entries

98146272:
cDNA cloning of a novel trypsin inhibitor with similarity to
pathogenesis-related proteins, and its
frequent expression in human brain cancer cells.

Peptide information for frame 1

ORF from 205 bp to 1695 bp; peptide length: 497
Category: strong similarity to known protein

```

1 MSCVLGGVIP LGLFLVCGS QGYLLPNVTI LEELLSKYQH NESHRSVRRA
51 IPREDKEEIL MLHNKLRGQV QPOASNMEYM TWDELEKSA AAWASQCIWE
101 HGPTSLVLSI GQNLGAHWGR YRSPGFHVQS WYDEVKDYTY PYPSECNPWC
151 PERCSGPMCT HYTQIVWATT NKIGCAVNTC RKMTVWGEVW ENAVYFVCNY
201 SPKGNWIGEA PYKNGRPCSE CPPSYGGSCR NNLCYREETY TPKPETDEMN
251 EVETAPIPEE NHVWLQPRVM RPTKPKKTSV VNYMTQVVRD DTKMKDRCKG
301 STCNRYQCPA GCLNHKAKIF GTLFYESSSS ICRAAIHYGI LDDKGLVDI
351 TRNGKVFFV KSERHGVQSL SKYKPSSEFM VSKVKVQDLD CYTTVAQLCP
401 FEKPATHCPR IHCPAHCKDE PSYWAPVFGT NIYADTSSIC KTAVHAGVIS

```

451 NESGGDQDVM PVDKKKTYVG SLRNGVQSES LGTPRDGKAF RIFAVRQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4b4, frame 1

TREMBLNEW:AF109674_1 gene: "Lgll"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgll) mRNA, complete cds., N = 1, Score = 968, P = 1.9e-97

TREMBL:D45027_1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds., N = 1, Score = 738, P = 4.5e-73

TREMBL:AB009609_1 gene: "HrTT-1"; Halocynthia roretzi HrTT-1 mRNA, complete cds., N = 1, Score = 345, P = 2e-31

PIR:JC5308 testis-specific, vespid, and pathogenesis-related protein 1 precursor - human, N = 1, Score = 337, P = 1.7e-30

>TREMBLNEW:AF109674_1 gene: "Lgll"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgll) mRNA, complete cds.

Length = 188

HSPs:

Score = 968 (145.2 bits), Expect = 1.9e-97, P = 1.9e-97
Identities = 160/185 (86%), Positives = 170/185 (91%)

Query: 61 MLHNKLRGQVQPOASNMEYMTWDDLEKSAAAWASQCIWEHGFTSLLVSIGQNLGAHWGR 120
MLHNKLRGQV P ASNMEYMTWD+ELE+SAAAWA +C+WEHGP SLLVSIGQNL HWGR
Sbjct: 1 MLHNKLRGQVPPASNMEYMTWDEELERSAAAWAQRLWEHGFPASLLVSIGQNLAVHWGR 60

Query: 121 YRSPGFHVQSWYDEVKDYTYPPSECNPWCPERCSPGPMCTHYTQIVWATTNKIGCAVNTC 180
YRSPGFHVQSWYDEVKDYTYPP ECNPWCPCRCSG MCTHYTQ+VWATTNKIGCAV+TC
Sbjct: 61 YRSPGFHVQSWYDEVKDYTYPPHECNPWCPERCSGAMCTHYTQMWWATTNKIGCAVHTC 120

Query: 181 RKMTVWGEVWENAVYFVCNYSKPGNWIAGEAPYKNGRPCSECPSPSYGGSCRNNLCYREETY 240
R M+VWG++WENAVY VCNYSKPGNWIAGEAPYK+GRPCSECP SYGG CRNNLCYREE Y
Sbjct: 121 RMSVWGDWENAVYLCNYSKPGNWIAGEAPYKHGRPCSECPSSYGGGCRNNLCYREEHY 180

Query: 241 TPKPE 245
KPE
Sbjct: 181 HQKPE 185

Pedant information for DKFZphtes3_4b4, frame 1

Report for DKFZphtes3_4b4.1

[LENGTH] 497
[MW] 55920.00
[pI] 8.36
[HOMOL] TREMBL:D45027_1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds. 6e-78
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YJL078c] 8e-12
[BLOCKS] BL01009E Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009D Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009C Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009A Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[PIRKW] glycoprotein 5e-22
[PIRKW] blocked amino end 5e-13
[PIRKW] brain 9e-30
[PIRKW] hydrolase 4e-09
[PIRKW] hemolymph coagulation 4e-09
[PIRKW] zymogen 4e-09
[PIRKW] alternative splicing 4e-09
[PIRKW] sperm 5e-22
[PIRKW] viroid-induced protein 2e-11
[PIRKW] venom 6e-18
[PIRKW] pyroglutamic acid 2e-11
[PIRKW] transmembrane protein 2e-10
[PIRKW] serine proteinase 4e-09
[SUPFAM] C-type lectin homology 4e-09
[SUPFAM] trypsin homology 4e-09

[SUPFAM] complement factor H repeat homology 4e-09
 [SUPFAM] cysteine-rich secretory protein 1 6e-24
 [SUPFAM] pathogenesis-related leaf protein 7e-15
 [PROSITE] MYRISTYL 8
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 3
 [PROSITE] SCP_AG5_PRI_SC7_2 1
 [PFAM] SCP-like extracellular Proteins
 [KW] All_Beta
 [KW] SIGNAL_PEPTIDE 23
 [KW] LOW_COMPLEXITY 1.21 %

SEQ MSCVLGGVPLGLLFLVCGSQGYLLPNVTILLEELLSKYQHNEHSRVRRAIPREDKEEIL
 SEGxxxxxx.....
 PRD ccc

SEQ MLHNKLRGQVQPQASNMEYMTWDDLEKSAAAWASQCIWEHGPTSLLVSIGQNLGAHWGR
 SEG
 PRD hhhhhhhcc

SEQ YRSPGFHVQSWYDEVKDYTYYPSECNPWCPCRCGPMCTHYTQIVWATTNKIGCAVNTC
 SEG
 PRD ccc

SEQ RKMTVWGEVWENAVYFVCNYSKGNWIGEAPYKNGRPCSECPSPSYGGSCRNNLCYREETY
 SEG
 PRD ccc

SEQ TPKPETDEMNEVETAPIEENHVWLQPRVMPRTKPKKTSAVNYMTQVVRCDTKMKDRCKG
 SEG
 PRD ccc

SEQ STCNRYQCPAGCLNHKAKIFGTLFYESSSSICRAAIHYGILDDKGGGLVDITRNGKVPFFV
 SEG
 PRD ccc

SEQ KSERHGVQSLSKYKPSSSFVSKVKVQDLDCYTTVAQLCPFEPATHCPRIHCPAHCKDE
 SEG
 PRD ecc

SEQ PSYWAPVFGTNIYADTSSICKTAVHAGVISNESGGDGDVMPVDKKKTYVGSRLRNGVQSES
 SEG
 PRD ccc

SEQ LGTPRDGKAFRIFAVRQ
 SEG
 PRD ccccccccccccccccc

Prosites for DKF2phtes3_4b4.1

PS00001	27->31	ASN_GLYCOSYLATION	PDOC00001
PS00001	41->45	ASN_GLYCOSYLATION	PDOC00001
PS00001	451->455	ASN_GLYCOSYLATION	PDOC00001
PS00004	181->185	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	276->280	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	464->468	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	179->182	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	362->365	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	483->486	PKC_PHOSPHO_SITE	PDOC00005
PS00006	29->33	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	81->85	CK2_PHOSPHO_SITE	PDOC00006
PS00006	130->134	CK2_PHOSPHO_SITE	PDOC00006
PS00006	453->457	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00007	385->393	TYR_PHOSPHO_SITE	PDOC00007
PS00008	111->117	MYRISTYL	PDOC00008
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	174->180	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

PS00008	227->233	MYRISTYL	PDOC00008
PS00008	300->306	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008
PS00008	470->476	MYRISTYL	PDOC00008
PS01010	195->207	SCP_AG5_PRI_SC7_2	PDOC00772

Pfam for DKFZphtes3_4b4.1

HMM_NAME	SCP-like extracellular Proteins		
HMM	*PQDEQDEWLKHNDFRQQVGRGLETRGNPGFPQPPAsNMnPMVWNDELAt		
	P + ++E+L HN +R QV	P ASNM M+W+DEL +	
Query	52	PREDKEEILMLHNKLRGQVQ-----PQASNMEYMTWDELEK	88
HMM	IAQnWANQCiFDHHDCCWNHsnYPYQONIAWWSsTAnnPWnWssMIQMwy		
	A WA+QCI +H ++ + S	GQN+ + + + + + +Q+WY	
Query	89	SAAAWASQCIWEHGPTSLLVSI---GQNLGAHWG---RYRSPGFHVQSWY	132
HMM	NEvkdYNYNWNTckGG.....NNFmVCGHYTQMvWRntfrIGCGRYICYC		
	+EVKDY Y +	+ +C HYTQ+VW+ T +IGC+ C+	
Query	133	DEVKDYTYPPSECNPWCPCRCGPMCTHYTQIVWATTNKIGCAVNTCRK	182
HMM	NNNWrkPDPWKhkWYYVCNYCPpGNYmN*		
	+ W + W+ +Y VCNy P+GN+++		
Query	183	MTVW--GEVWENAVYFVCNYSPKGNWIG	208

DKFZphtes3_4f17

group: testes derived

DKFZphtes3_4f17 encodes a novel 656 amino acid protein with weak similarity to methyl-CpG-binding proteins.

Methylation at the DNA sequence 5'-CpG is required for mammalian development. Methyl-CpG-binding proteins bind specifically to methylated DNA via a related amino acid motif and can repress transcription. The novel protein does not contain such a motive. No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to methyl-CpG-binding protein

extension of HS557771/HS278337,
there are some differences to these sequences

Sequenced by AGOWA

Locus: /map="18"

Insert length: 2320 bp

Poly A stretch at pos. 2266, polyadenylation signal at pos. 2251

```

1  GGCAGGTTTCG CGGGTCGCTG GCGGGGGTTCG TGAGGGAGTG CGCCGGGAGC
51  GGAGATATGG AGGGAGATGG TTCAGACCCA GAGCCTCCAG ATGCCGGGGA
101  GGACAGCAAG TCCGAGAATG GGGAGAATGC GCCCATCTAC TGCATCTGCC
151  GCAAACCGGA CATCAACTGC TTCATGATCG GGTGTGACAA CTGCAATGAG
201  TGGTTCCATG GGGACTGCAT CCGGATCACT GAGAAGATGG CCAAGGCCAT
251  CCGGGAGTGG TACTGTCGGG AGTGCAGAGA GAAAGACCCC AAGCTAGAGA
301  TTCGCTATCG GCACAAGAAG TCACGGGAGC GGGATGGCAA TGAGCGGGAC
351  AGCAGTGAGC CCGGGATGTA GGGTGGAGGG CGCAAGAGGC CTGTCCTTGA
401  TCCAGACCTG CAGCGCCGGG CAGGGTCAGG GACAGGGGTT GGGGCCATGC
451  TTGCTCGGGG CTCTGCTTCG CCCACAAAT CCTCTCCGA GCCCTTGGTG
501  GCCACACCCA GCCAGCATCA CCAGCAGCAG CAGCAGCAGA TCAAACGGTC
551  AGCCCGCATG TGTGGTGAGT GTGAGGCATG TCGGCCCACT GAGGACTGTG
601  GTCACGTGTA TTTCTGTCGG GACATGAAGA AGTTCGGGGG CCCCAACAAG
651  ATCCGGCAGA AGTGCCGGCT GCGCCAGTGC CAGCTGCGGG CCCGGGAATC
701  GTACAAGTAC TTCCCTTCCT CGCTCTCACC AGTGACGCC TCAGAGTCCC
751  TGCCAAGGCC CCGCCGGCCA CTGCCACCC AACAGCAGCC ACAGCCATCA
801  CAGAAGTTAG GGGCATCCG TGAAGATGAG GGGGCAGTGG CGTCATCAAC
851  AGTCAAGGAG CCTCCTGAGG CTACAGCCAC ACCTGAGCCA CTCTCAGATG
901  AGGACCTACC TCTGGATCCT GACCTGTATC AGGACTTCTG TGCAGGGGCC
951  TTTGATGACC ATGGCCTGCC CTGGATGAGC GACACAGAAG AGTCCCATT
1001  CCTGGACCCC GCGCTGCGGA AGAGGGCAGT GAAAGTGAAG CATGTGAAGC
1051  GTCGGGAGAA GAAGTCTGAG AAGAAGAAGG AGGAGCGATA CAAGCGGCAT
1101  CGGCAGAGAG AGAAGCACAA GGATAAATGG AAACACCCAG AGAGGGCTGA
1151  TGCCAAGGAC CCTGCGTCAC TGCCCCAGTG CCTGGGGCCC GGCTGTGTGC
1201  GCCCCGCCCA GCCCAGCTCC AAGTATTGCT CAGATGACTG TGGCATGAAG
1251  CTGGCAGCCA ACCGCATCTA CGAGATCCTC CCCCAGCGCA TCCAGCAGTG
1301  GCAGCAGAGC CCTTGCATTG CTGAAGAGCA CGCAAGAAG CTGCTCGAAC
1351  GCATTGCGCG AGAGCAGCAG AGTGCCCGCA CCGCCTTCA GGAATGGAA
1401  CGCCGATTCC ATGAGCTTGA GGCCATCATT CTACGTGCCA AGCAGCAGGC
1451  TGTGCGCAG GATGAGGAGA GCAACGAGGG TGACAGTGAT GACACAGACC
1501  TGCAGATCTT CTGTGTTTCC TGTGGGCACC CCATCAACCC ACGTGTGGCC
1551  TTGCGCCACA TGGAGCGCTG CTACGCCAAG TATGAGAGCC AGACGTCCTT
1601  TGGGTCCATG TACCCACAC GCATTGAAGG GGCCACACGA CTCTTCTGTG
1651  ATGTGTATAA TCCTCAGAGC AAAACATACT GTAAGCGGCT CCAGGTGCTG
1701  TGCCCCGAGC ACTCACGGGA CCCCAAAGTG CCAGCTGACG AGGTATGCGG
1751  GTGCCCCCTT GTACGTGATG TCTTTGAGCT CACGGGTGAC TTCTGCCGCC
1801  TGCCCAAGCG CCAGTGCAAT CGCCATTACT GCTGGGAGAA GCTGCGGCGT
1851  GCGGAAGTGG ACTTGGAGCG CGTGCCTGTG TGGTACAAGC TGGACGAGCT
1901  GTTTGAGCAG GAGCGCAATG TGCGCACAGC CATGACAAAC CGCGCGGGAT
1951  TGCTGGCCCT GATGCTGCAC CAGACGATCC AGCACGATCC CCTCACTACC
2001  GACCTGCGCT CCAGTGCCGA CCGCTGAGCC TCCTGGCCCG GACCCCTTAC
2051  ACCCTGCATT CCAGATGGGG GAGCCGCGCG GTGCCCGTGT GTCCGTTTCT
2101  CCACTCATCT GTTCTCCGG TTCTCCCTGT GCCCATCCAC CGGTTGACCG
2151  CCCATCTGCC TTTATCAGAG GGAAGTGTCC CGTCGACATG TTCAGTGCCT
2201  GGTGGGGCTG CGGAGTCCAC TCATCCTTGC CTCTCTTCCC TGGGTTTTGT
2251  TAATAAAATT TTGAAGAAAC CAAAAAATAA AAAAAAATAA AAAAAAATAA
2301  AAAAAAATAA AAAAAAATAA

```

BLAST Results

 Entry HS557771 from database EMBLEST:
 Human chromosome 18 clone 2 mRNA sequence.
 Score = 7582, P = 0.0e+00, identities = 1560/1598

Entry HS278337 from database EMBLEST:
 H.sapiens mRNA, expressed sequence tag ICRFp507H02194 (5')
 Score = 6339, P = 9.0e-281, identities = 1307/1347

Entry HS095149 from database EMBL:
 human STS WI-6941.
 Score = 1210, P = 2.2e-49, identities = 246/251

Medline entries

98449942:
 Identification and characterization of a family of mammalian methyl-CpG binding proteins.

9824997:
 Gene silencing by methyl-CpG-binding proteins.

Peptide information for frame 3

ORF from 57 bp to 2024 bp; peptide length: 656
 Category: similarity to known protein

```

1 MEGDGSDEP PDAGEDSKSE NGENAPIYCI CRKPDINCFM IGCDCNCNEWF
51 HGDCIRITEK MAKAIREWYC RECREKDPKL EIRYRHKKS ERDGNERNDS
101 EPRDEGGGRK RPVPDPDLQR RAGSGTGVA MLARGSASPH KSSPQPLVAT
151 PSQHHQQQQQ QIKRSARMCG ECEACRRTD CGHCDFCRDM KKGFGPNKIR
201 QKCRRLRQQL RARESYKYFP SLSPTVPSE SLPRPRRPLP TQQQPQPSQK
251 LGRIREDEGA VASSTVKEPP EATATPEPLS DEDLPLDPL YQDFCAGAFD
301 DHGLPWMSDT EESPFDPAL RKRAVKVKHV KRREKKSEKK KEERYKRHRQ
351 KQKHDKWKH PERADAKDPA SLPQCLGPGC VRPAQPSKY CSDDCGMKLA
401 ANRIYEILPQ RIQQWQSPC IAEHGGKLL ERIRREQQA RTRLQEMERR
451 FHELEAILR AKQAVREDE ESNEGSDDT DLQIFCVSCG HPINPRVALR
501 HMERCYAKYE SQTSGSMYP TRIEGATRLF CDVYNPQSKT YCKRLQVLCF
551 EHSRDPKVPA DEVCGPLVR DVFELTGDFC RLPKRQCNRH YCWEKLRAE
601 VDLEVRVWY KLDELFEQER NVRTAMTNRA GLLALMLHQT IQHDLTTDL
651 RSSADR
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4f17, frame 3

TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11, N = 2, Score = 316, P = 8.8e-27

TREMBL:HSAB2331_1 gene: "KIAA0333"; Human mRNA for KIAA0333 gene, partial cds., N = 2, Score = 163, P = 2.8e-13

TREMBL:SPCC594_5 gene: "SPCC594.05c"; product: "putative transcriptional regulatory protein, phd finger containing"; S.pombe chromosome III cosmid c594., N = 3, Score = 168, P = 3.6e-12

TREMBL:AF072240_1 gene: "Mbd1"; product: "methyl-CpG binding protein MBD1"; Mus musculus methyl-CpG binding protein MBD1 (Mbd1) mRNA, complete cds., N = 2, Score = 189, P = 7.6e-11

>TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11
 Length = 523

HSPs:

Score = 316 (47.4 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
 Identities = 100/336 (29%), Positives = 167/336 (49%)

Report for DKFZphtes3_4f17.3

SEQ RAGSGTGVGAMLRGASASP HKSSPOPLVATPSQHQQQQQIKRSARMCGECEACRRTED
 SEGxxxxxxxxx.....
 PRD cccccccceeeccccccccccccccccchhhhhhhhhhhhhhhhhccccccccccccc
 COILS

Prosites for DKF2phtes3_4f17.3

(No Pfam data available for DKFZphtes3_4f17.3)

DKFZphtes3_4f5

group: signal transduction

DKFZphtes3_4f5 encodes a novel 790 amino acid protein similar to beta-transducins.

The protein contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a Cytochrome C family heme-binding site signature is present. The protein is larger (790 amino acids) than the usual eukaryotic G-beta transducins (about 340 amino acids).

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to *S.pombe* "beta-transducin"

complete cDNA, EST hits

complete cds,

on genomic level encoded by HS313D11, at least 7 exons these exons

match

only partially with the predicted transcripts in HS313D11

Sequenced by AGOWA

Locus: /map="16p13.3"

Insert length: 3166 bp

No poly A stretch found, no polyadenylation signal found

```

1  GCGCGCTTCC GCGCGGCGG TTCCGGACAA CCGTGCGCTT TTAGTAAAAG
51  ATTTGGGGTTC GCGCGGGGGA GAAGGGCTGC CCCGGGCCCT CTGGTTCTCG
101 TCCCGCAGCG TCCGCTCCCC CGCGCCACTG CGCGCTCCC AGGAACCCCTG
151 TACTCCGGGG TCGCCGGCTT CTCTCCTGCC TCCGGTCCCG CCAGACACCT
201 CGAGCTCCTT AAGTAGCTCG GTCCTTGACG TCCCTCTGGG CCCTTCCCGC
251 GTCTATCGCC TGAGTCCCGG GGCCCTCTA GCCCTCTGTT CCCTCCCTC
301 TTTTGTTCCT CCCTAGAGCC CGCCCGCCCT CAGGGCTGAC AGTGTGGACG
351 GCGGGAGTCT CCTCGCTCCC CTGCTGGGAT TGAAGTACCG AGCGTTTAGT
401 GACTGCCCAG ATCTGGCTGA TGGGGGTACC GAGAGGTGGC CTGGGCGGGG
451 AATGTCCAGC TAGAGTCTTC CGTGGAAGTC AGACATGAAA CTGACAGGCC
501 TAAGGGAAGC TAGGAAGTCC CCTCACCGCT CAGCCAGGGT GATGGGCTGG
551 ACTGACAGAC TCCAGTGAAT TTGAGCTTGC CTGTGAGGCT GATTGGCTGA
601 TAGACAGCCC TGGATTGGCT CACTAAGACT GACCAGCCCG GGACCAAGCA
651 GTTCTGGGGT CCCAACCTGG GTGGAAGGTC TGAAGTGATG ACCCACCCAG
701 GCTGACCAGG CCAGCCACAC TCACTGACCT CCTGACCCCT GACCTCATCA
751 CCTGTGACAG CATGGAGAAG ATGTCCCGTG TGACCACAGC CTGGGTGGC
801 AGCGTGCTGA CAGGCGGCAC CATGCACTGC CACCTGGATG CTCCCGCCAA
851 TGCCATCAGT GTGTGCCGCG ACGCAGCCCA GGTGGTCTGT GCAGGCCGTA
901 GCATCTTCAA GATCTATGCC ATCGAGGAGG AACAGTTCGT GGAAAAGCTG
951 AACCTGCGTG TGGGGCGCAA GCCTTCGCTT AACCTGAGCT GTGCTGACGT
1001 GGTCTGGCAC CAGATGGATG AGAACCTGCT GGCCACAGCA GCCACCAATG
1051 GCGTGGTGGT CACGTGGAAC CTGGGCGGCG CATCCCGCAA CAAGCAGGAC
1101 CAGTGTTCAC CAGAACACAA GCGCAGCGTA AACAAAGTCT GCTTCCACCC
1151 CACCGAAGCC CACGTGCTGC TCAGTGGCTC CCAGGATGGC TTCATGAAGT
1201 GCTTTGACCT CCGCAGAAAG GACTCTGTCA GCACCTTCTC GGGCCAGTCG
1251 GAGAGCGTGC GGGACGTGCA GTTCAGTATC CGGGACTACT TCACCTTCGC
1301 CTCACCTTT GAGAACGGCA ATGTGCAGCT CTGGGACATC CGGCGTCCCG
1351 ACCGGTGCGA GAGGATGTTC ACAGCCACCA ACGGACCCGT CTCTGCTGC
1401 GACTGGCACC CCGAGGACAG GGGCTGGTTG GCCACTGGAG GGCGCGACAA
1451 GATGGTGAAG GTCTGGGACA TGACCACGCA CCGTGCCAAG GAGATGCACT
1501 GTGTGCAGAC CATCGCTCG GTGGCCCGTG TGAAGTGGCG GCCAGAGTGC
1551 CGCCACCACC TGGCCACGTG CTCCATGATG GTGGACCACA ACATCTATGT
1601 TTGGGACGTG CGCCGGCCCT TCGTGCCAGC TGCCATGTTT GAGGAACACC
1651 GAGACGTAC CACGGGAATT GCCTGGCGCC ACCCCACGCA CCCCTCCTTC
1701 CTGCTGTCTG GCTCCAAGGA CAGCTCGCTG TGCCAGCACC TGTTCGCGA
1751 CGCCAGCGAG CCCGTCGAGC CGCGCAACCC TGAGGGCCTC TGCTACGGCC
1801 TCTTCGGGGA CCTGGCCTTC GCCGCCAAGG AGAGCCTCGT GGCTGCCGAG
1851 TCGGGGCGCA AGCCCTACAC TGGCGACCGG CGCCACCCCA TCTTCTTTAA
1901 GCGCAAGCTG GACCCCTGCC AGCCCTTCGC AGGCCTCGCC TCCAGTGCCC
1951 TCAAGTGTCT TGAGACGGAG CCAGGTGGCG GCGGCATGCG CTGGTTTGTG
2001 GACACAGCTG AGCGTTATGC GCTGGCTGGC CGGCCACTGG CCGAGCTCTG
2051 TGACCACAAC GCAAAGGTGG CTCGAGAGCT TGGCCGCAAC CAGGTGGCGC
2101 AAACGTGGAG CATGTGCGG ATCATCTACT GCAGCCCTGG CTAAGTGCCC
2151 ACTGCAAAAC TCAACCACAG TGTGGGCAAG GGTGGCTCCT GTGGCCTCCC
2201 GCTCATGAAC AGTTTCAACC TGAAGGATAT GGCCCCAGGG TTGGGCAGTG
2251 AGACGCGGCT GGACCGCAGC AAAGGAGATG CACGGAGCGA CACAGTTCTG
2301 CTCGACTCCT CGGCCACACT CATACCAAT GAGGATAACG AGGAAACCGA
2351 GGGCAGCGAC GTACCTGCGG ACTACCTGCT GGGTGACGTG GAAGGTGAGG

```

```

2401 AGGACGAGCT GTACCTGCTG GATCCGGAAC ACGCGCACCC CGAGGACCCT
2451 GAGTGCGTGC TGCCGCAGGA GGCCTTTCCG CTGCGCCACG AGATCGTGGA
2501 CACGCCCTCC GGACCCGAGC ACCTGCAGGA CAAGGCCGAC TCCCCGCACG
2551 TGAGCGGCAG CGAGGCGGAT GTGGCCTCCC TGGCCCCCGT GGACTCCTCC
2601 TTCTCGCTCC TGTCTGTCTC ACACGCGCTC TACGACAGCC GCCTGCCGCC
2651 CCACTTCTTC GCGTGTCTGG TGCGCGACAT GCTGCACTTC TACGCTGAGC
2701 AGGGCGACGT GCAGATGGCT GTGTCTGTGC TCATCGTCCT GGGTGAACGG
2751 GTGCGCAAGG ACATCGACGA GCAGACCCAG GAGCACTGGT ACACCTCCTA
2801 CATCGACCTG CTGCAGCGCT TCCGCCTCTG GAACGTGTCC AACGAGGTGG
2851 TCAAGCTGAG CACCAGCCGC GCCGTCAGCT GCCTCAACCA GGCCTCCACC
2901 ACCCTGCACG TCAACTGCAG CCACTGCAAG CGGCCCATGA GCAGCCGGGG
2951 CTGGGTCTGC GACAGGTGCC ACCGCTGCGC CAGCATGTGT GCCGTCTGCC
3001 ACCACGTAGT CAAGGGTCTC TTCGTGTGGT GCCAGGGCTG CAGCCACGGC
3051 GGCCACCTGC AGCACATCAT GAAATGGCTG GAAGGCAGCT CCCACTGTCC
3101 CGCAGGCTGC GGCCACCTCT GCGAGTACTC CTGACGGGGC ATCTGTGGG
3151 CTTGCCCCGG CGGCCG

```

BLAST Results

Entry HS313D11 from database EMBL:
 Human DNA sequence from cosmid 313D11 from a contig on the short arm of
 chromosome 16. Contains ESTs, STS and CpG islands.
 Score = 6238, P = 0.0e+00, identities = 1318/1391

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 762 bp to 3131 bp; peptide length: 790
 Category: similarity to known protein

```

1 MEKMSRVTTA LGGSVLTGRT MHCHLDAPAN AISVCRDAAQ VVVAGRSIFK
51 IYAIIEEQFV EKLNLRVGRK PSLNLSCADV VWHQMDENLL ATAATNGVVV
101 TWNLGRPSRN KQDQLFTEHK RTVNKVCFHP TEAHVLLSGS QDGFMKCFDL
151 RRKDSVSTFS GQSESVRDVQ FSIRDYFTFA STFENGIVQL WDIRRPDRCE
201 RMFTAHNGPV FCCDWHPEDR GWLATGGRDK MVKVDWMTTH RAKEMHCVQT
251 IASVARVKWR PECRHHLATC SMMVDHNIYV WDVRRPFVPA AMFEEHRDVT
301 TGIARHHPHD PSFLLSGSKD SSLCOHLFRD ASQPVERANP EGLCYGLFGD
351 LAFAAKESLV AESGRKPYT GORRHPIFFK RKLDPAPFPA GLASSALSVE
401 ETEPGGGGMR WFDVTAERYA LAGRPLAELC DHNAKVAREL GRNQVAQTWT
451 MLRIIYCSPG LVPTANLNHS VGKGGSCGLP LMNSFNLKDM APGLGSETRL
501 DRSKGDARS DTVLLDSSATL ITNEDNEETE GSDVPADYLL GDVEGEDEL
551 YLLDPEHAHP EDPECVLPQE AFPLRHEIVD TPPGPEHLQD KADSPHVSQS
601 EADVASLAPV DSSFSLLSVS HALYDSRLPP DFFGVLVDRM LHFYAEQGDV
651 QMAVSVLIVL GERVRKDIDE QTQEHWYTSY IDLLQRFRLW NVSNEVVKLS
701 TSRAVSCLNQ ASTTLHVNCN HCKRPMSSRG WVCDCRHRCA SMCACVHHVV
751 KGLFVWCQGC SHGGHLQHIM KWLEGSSHCP AGCGHLCEYS

```

BLASTP hits

Entry YDSB SCHPO from database SWISSPROT:
 HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN C4F8.11 IN
 CHROMOSOME I. >TREMBL:SPAC4F8_11 gene: "SPAC4F8.11"; product:
 "beta-transducin"; S.pombe chromosome I cosmid c4F8.
 Score = 404, P = 3.0e-42, identities = 169/639, positives = 278/639

Entry PEX7 HUMAN from database SWISSPROT:
 PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
 >TREMBL:HSU76560_1 gene: "Pex7"; product: "peroxisome targeting signal
 2 receptor"; Human peroxisome targeting signal 2 receptor (Pex7) mRNA,
 complete cds. >TREMBL:HSU88871_1 gene: "HsPEX7"; product: "HsPex7p";
 Human HsPex7p (HsPEX7) mRNA, complete cds.
 Score = 220, P = 1.1e-15, identities = 62/244, positives = 107/244

Entry PEX7 MOUSE from database SWISSPROT:
 PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
 >TREMBL:MMU69171_1 product: "peroxisomal PTS2 receptor"; Mus musculus
 peroxisomal PTS2 receptor mRNA, complete cds.
 Score = 214, P = 5.3e-15, identities = 60/240, positives = 106/240

Entry ATAC2294.7 from database TREMBL:
 gene: "Flp17.7"; Arabidopsis thaliana chromosome I BAC Flp17 genomic
 sequence, complete sequence.
 Score = 232, P = 3.4e-14, identities = 68/260, positives = 120/260

Entry S66835 from database PIR:
 probable membrane protein YOL138c - yeast (*Saccharomyces cerevisiae*)
 >TREMBL:SCYOL138C_1 *S.cerevisiae* chromosome XV reading frame ORF
 YOL138c
 Score = 136, P = 2.5e-13, identities = 24/77, positives = 44/77

Alert BLASTP hits for DKFZphtes3_4f5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_4f5, frame 3

Report for DKFZphtes3_4f5.3

```
[LENGTH]      790
[MW]           88207.10
[pI]           6.05
[HOMOL]        SWISSPROT:YDSB SCHPO HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN
C4F8.11 IN CHROMOSOME I. 9e-44
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YOL138c] 5e-16
[FUNCAT]       10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YCR072c beta-transducin family]
3e-10
[FUNCAT]       04.05.01.01 general transcription activities [S. cerevisiae, YBR198c
TAF90 - TFIID subunit] 9e-09
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YLL011w] 1e-07
[FUNCAT]       30.09 organization of intracellular transport vesicles [S. cerevisiae,
YDL195w] 2e-07
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL195w]
2e-07
[FUNCAT]       30.19 peroxisomal organization [S. cerevisiae, YDR142c] 4e-07
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c]
4e-07
[FUNCAT]       08.10 peroxisomal transport [S. cerevisiae, YDR142c] 4e-07
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YER107c] 4e-07
[FUNCAT]       04.07 rna transport [S. cerevisiae, YER107c] 4e-07
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YER107c] 4e-07
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 5e-07
[FUNCAT]       06.13 proteolysis [S. cerevisiae, YGL003c] 5e-07
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 8e-07
[FUNCAT]       04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-06
[FUNCAT]       03.13 meiosis [S. cerevisiae, YLR129w] 3e-06
[FUNCAT]       03.25 cytokinesis [S. cerevisiae, YCR057c] 1e-05
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c]
1e-05
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 2e-04
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YOR272w] 6e-04
[SCOP]         d1gothb_2.46.3.1.1 beta1-subunit of the signal-transducing 5e-06
[PIRKW]        duplication 7e-10
[PIRKW]        signal transduction 7e-08
[PIRKW]        peroxisome 9e-06
[PIRKW]        heterotrimer 7e-08
[PIRKW]        GTP binding 7e-08
[PIRKW]        peroxisome biogenesis 9e-06
[PIRKW]        transmembrane protein 1e-14
[SUPFAM]       MSI1 protein 7e-10
[SUPFAM]       WD repeat homology 1e-14
[SUPFAM]       GTP-binding regulatory protein beta chain 7e-08
[SUPFAM]       PRL1 protein 3e-08
[SUPFAM]       coatamer complex beta' chain 1e-06
[PROSITE]      CYTOCHROME_C 1
[PROSITE]      WD_REPEATS 3
[PROSITE]      MYRISTYL 10
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 2
[PROSITE]      CK2_PHOSPHO_SITE 11
```

[PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 7
 [PROSITE] ASN_GLYCOSYLATION 4
 [PFAM] WD domain, G-beta repeats
 [KW] All_Beta
 [KW] 3D
 [KW] LOW_COMPLEXITY 2.28 %

```

SEQ      MEKMSRVTTALGGSVLTGRMHCHLDAPANAISVCRDAAQVVVAGRSIFKIYAIEEEQFV
SEG      .....
lgotB    .....

SEQ      EKLNLRVGRKPSLNLSCADVWHQMDENLLATAATNGVVVTWNLGRPSRNKQDQLFTEHK
SEG      .....
lgotB    .....TTCEEEEEETTTTEEEET-TTCEEE--EEECCE

SEQ      RTVNKVCFHPTAEHVLLSGSQDGMKCFDLRRKDSVSTFSGQSESVRDVQFSIRDYFTFA
SEG      .....
lgotB    CCEEEEEETTT-TCEEEEEETTTTEEEETTTTEEEECBTTCCEEEEEETTTTTEEE

SEQ      STFENGVLWDIIRRPDRCEMFTAANGPVFCCDWHPEDRGWLATGGRDKMVKVWDMTTH
SEG      .....
lgotB    E-ETTTTEEEETTTTTEEE-EEECCECEEEEEE-TTTTCCEEEETTTTEEEEC....

SEQ      RAKEMHCVQTIASVARVKWRPECRHHLATCSMMVDHNIYVWDVRRPFVPAAMFEEHRDVT
SEG      .....
lgotB    .....

SEQ      TGIARHPHDPFSLSGSKDSSLQHLFRDASQPVERANPEGLCYGLFGLAFAAKESLV
SEG      .....
lgotB    .....

SEQ      AAESGRKPYTGDRRHPIFFKRKLDPAEPFAGLASSALSVFETEPGGGMRWFVDTAERYA
SEG      .....
lgotB    .....

SEQ      LAGRPLAELCDHNAKVARELGRNQVAQTWMLRIIYCSPLVPTANLNHSVKGKGGSCGLP
SEG      .....
lgotB    .....

SEQ      LMNSFNLKDMAPGLGSETRLDRSKGDARSDTVLLDSSATLITNEDNEETEGSDVPADYLL
SEG      .....xxxx
lgotB    .....

SEQ      GDVEGEDELYLLDPEHAHPEDPECVLPQEAFLRHEIVDTPPGPEHLQDKADSPHVSQS
SEG      xxxxxxxxxxxxxxxx
lgotB    .....

SEQ      EADVASLAPVDSSFSLLSVSHALYDSRLPPDFFGVLRDMLHFYAEQGDVQMAVSVLIVL
SEG      .....
lgotB    .....

SEQ      GERVRKDIDEQTQEHWYTSYIDLLQRFRLWNVSNEVVKLSTSRVAVSCLNQASTTLHVNC
SEG      .....
lgotB    .....

SEQ      HCKRPMSSRGVWCDRCHRCASMCVCHHVVKGLFVWCQGC SHGHLQHIMKWLEGSSHCP
SEG      .....
lgotB    .....

SEQ      AGCGHLCEYS
SEG      .....
lgotB    .....

```

Prosites for DKFZphtes3_4f5.3

PS00001	74->78	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	691->695	ASN_GLYCOSYLATION	PDOC00001
PS00001	718->722	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	152->156	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	165->168	PKC_PHOSPHO_SITE	PDOC00005
PS00005	172->175	PKC_PHOSPHO_SITE	PDOC00005
PS00005	239->242	PKC_PHOSPHO_SITE	PDOC00005
PS00005	364->367	PKC_PHOSPHO_SITE	PDOC00005
PS00005	701->704	PKC_PHOSPHO_SITE	PDOC00005

PS00005	727->730	PKC_PHOSPHO_SITE	PDOC00005
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	398->402	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	522->526	CK2_PHOSPHO_SITE	PDOC00006
PS00006	598->602	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	679->683	CK2_PHOSPHO_SITE	PDOC00006
PS00007	337->346	TYR_PHOSPHO_SITE	PDOC00007
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	97->103	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	161->167	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	342->348	MYRISTYL	PDOC00008
PS00008	391->397	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	474->480	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009
PS00009	364->368	AMIDATION	PDOC00009
PS00190	743->749	CYTOCHROME_C	PDOC00169
PS00678	90->105	WD_REPEATS	PDOC00574
PS00678	223->238	WD_REPEATS	PDOC00574
PS00678	269->284	WD_REPEATS	PDOC00574

Pfam for DKFZphtes3_4f5.3

HMM_NAME	WD domain, G-beta repeats		
HMM	*MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*		
	++ HN++V C+ ++P+ R +++G++D+ +++WD		
Query	203	FTAHNGP VFCCDWH PEDRGWLATGGRDKMVKVWD	236

DKFZphtes3_4h6

group: intracellular transport/trafficking

DKFZphtes3_4h6 encodes a novel 622 amino acid protein with strong similarity to the kinesin light chain.

Kinesin is a microtubule-based motor protein that pulls vesicles or organelles towards the plus end of microtubules. Structural changes in the protein that drive motility are coupled to ATP binding and hydrolysis. The novel protein is similar to kinesin light chain, which is part of the functional kinesin holoenzyme tetrameric protein. The light chain has been proposed to function in coupling of cargo to the heavy chain or in the modulation of the ATPase activity of the heavy chain. The novel protein contains two kinesin light chain repeats and one RGD cell-attachment site.

The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

strong similarity to Kinesin light chain

complete cDNA, complete cds, start at 150, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 2992 bp

Poly A stretch at pos. 2914, polyadenylation signal at pos. 2893

```

1  GCGGGGATGC AGGCGGCGGG ACCGGCTCGC GGGTGCGGGT CCGGGTGAAG
51  CCGGAGGCAG CCAGAGTCGG AGCCGGGCCC GAGCACCAGG CGCAGGCCCC
101 GCGCCCGCCT GCCCGCACCC TCGTCCTCAC AGACGCCACA GCCATGGCCA
151 TGATGGTGTT TCCGCGGGAG GAGAAGCTGA GCCAGGATGA GATCGTGCTG
201 GGCACCAAGG CTGTCATCCA GGGACTGGAG ACTCTGCGTG GGGAGCATCG
251 TGCCCTGCTG GCTCCTCTGG TTGCACCTGA GGCCGGCGAA GCCGAGCCTG
301 GCTCGCAGGA GCGCTGCATC CTCCTGCGTC GCTCCCTGGA AGCCATTGAG
351 CTTGGGCTGG GGGAGGCCCA GGTGATCTTG GCATTGTCTG GCCACCTGGG
401 GGCTGTAGAA TCAGAGAAGC AGAAGCTGCG GCGCGAGGTG CGGCGTCTGG
451 TGCAGGAGAA CCACTGGCTG CGTGAGGAGC TGGCGGGGAC ACAGCAGAAG
501 CTGCAGCGCA GTGAGCAGGC CGTGGCCAGG CTCGAGGAGG AGAAGCAGCA
551 CTTGCTGTTC ATGAGCCAGA TCCGCAAGTT GGATGAAGAC GCCTCCCTTA
601 ACGAGGAGAA GGGGGACGTC CCCAAAGACA CACTGGATGA CCTGTTCCCT
651 AATGAGGATG AGCAGAGCCC AGCCCTAGC CCAGGAGGAG GGGATGTGTC
701 TGGTCAGCAT GGGGGCTACG AGATCCCGGC CCGGCTCCGC ACCCTGCACA
751 ACCTGGTGAT CCAATACGCC TCACAGGGCC GCTACGAGGT AGCTGTGCCA
801 CTCTGCAAGC AGGCACTCGA AGACCTGGAG AAGACGTCAG GCCACGACCA
851 CCCTGACGTT GCCACCATGC TGAACATCCT GGCACCTGGT TATCGGGATC
901 AGAACAACTA CAAGGAGGCT GCCCACCTGC TCAATGATGC TCTGGCCATC
951 CCGGAGAAAA CACTGGGCAA GGACCACCCA GCCGTGGCTG CGACACTAAA
1001 CAACCTGGCA GTCTGTATG GCAAGAGGGG CAAGTACAAG GAGGCTGAGC
1051 CATGTGTCAA CGGGGCACTG GAGATCCGGG AGAAGGTCCT GGGCAAGTTT
1101 CACCCAGATG TGGCCAAGCA GCTCAGCAAC CTGGCCCTGC TGTGCCAGAA
1151 CCAGGGGAAA GCTGAGGAGG TTGGAATATTA CTATCGGCGG GCACCTGGAG
1201 TCTATGTCTA ACGCCTCGGG CCCGATGACC CCAATGTGGC CAAGACCAAG
1251 AACAACTTGG CTTCTGTGTA CCTGAAGCAG GGCAAGTACC AGGATGCGGA
1301 GACCTTGTAC AAGGAGATCC TCACCCGCGC TCATGAGAAA GAGTTTGGCT
1351 CTGTCAATGG GGACAACAAG CCCATCTGGA TGCACGAGA GGAGCGGGAG
1401 GAAAGCAAAG ATAAGCGCCG GGACAGCGCC CCCTATGGGG AATACGGCAG
1451 CTGGTACAAG GCCTGTAAAG TAGACAGCCC CACAGTCAAC ACCACCCTGC
1501 GCAGCTTGGG GGCCCTATAC CGGCGCCAGG GCAAGCTGGA AGCCCGGCAC
1551 AACTAGAGG ACTGTGCCAG CCGTAACCGC AAGCAGGGTT TGGACCCCGC
1601 AAGCCAGACC AAGTGCTAG AACTGCTGAA AGATGGCAGT GGCAGGCGGG
1651 GAGACCGCCG CAGCAGCCGA GACATGGCTG GGGGTGCCGG GCCTCGGTCT
1701 GAGTCTGACC TCGAGGACGT GGGACCTACA GCTGAGTGGA ATGGGGATGG
1751 CATTGGCTCC TTGAGGCGCA GCGGTTCCCT TGGGAAACTC CGGGATGCCC
1801 TGAGGCGCAG CAGTGAGATG CTGGTAAAGA AGCTGCAGGG GGGCACCCCC
1851 CAGGAGCCCC CTAACCCAG GATGAAGCGG GCCAGTTCCC TCAACTTCCT
1901 CAACAAGAGC GTGGAAGAGC CGACCCAGCC TGGAGGCACA GGTCTCTCTG
1951 ACAGCCGCAC TCTCAGCTCC AGCTCCATGG ACCTCTCCCG ACGAAGCTCC
2001 CTGTGTGGGT AATGCTGAAG GGGCAGCCAG TCACCAGAGC GCCACCTGG
2051 CACACCCCCC TCACCCAGC CCGTGCATG GGCCTGCTGC TTGTCCCGCC
2101 TGTCTCTCCC ACAGCCCTG TCTTTTCTGT TCAATCTCAG GGTAACCTTC
2151 TCCCTTGTC TCTCAGCTG AGCCCTGGAG GCTGGGCTG CCCACTCCAG
2201 CTCATCCTCT TATTATTCC TTCCAGCAGG GCCCTCTTCC CTAGGTTGCG
2251 GCCAGCAGGA GGTGCCGGCT GGAGTCTCCA CCATAGACTC AGTGGCCTGG
2301 CCTCCCCAGA CCCCAGAGCC AAGAACAATA AGCACTCGCC GGCCCTTCGG
2351 CACCTTCGCC CTCCTCCCG ACTCAACCCG GCGGTTGCTT CTGTATATAG
2401 AGAAATAAGT TATTGGCCGC GCGCTCCCT TCAGTCCACG GTACTACCCG

```



```

2451 GGCCTCCCTT CGTCCCTCTT CTAGTGGTAC CGCCCAGGCC TTAATCACCC
2501 CCATTCCGTG CGGTGGTATC TCCCAGGCTC TACATTCTCG GGAGCGGCGC
2551 CTCCCAAGGG GGTCTTGGGA CCTTCTCGCG CTCCTCCTGG CCTCTGAGGG
2601 ATGCGTCTTA CCCGCGCCAT CGCCCCGTGG CCCAGGACGG GGACCTCCCC
2651 TTAGTCCGTC CTCCCACCGC CGGGCCCTGC CCCGCATCCC GGCCTTATGC
2701 ACTGCCCTCT CCACCCGGCC CCGCCCAGGC ACGGCCGACC CCGCCCCGGG
2751 CACCGCCAC CGAGCCATCC TGCCTCGCCT CCCCCACGC CTGCAGCTTC
2801 TCGCGAGGGG CGGCGACGGT CCCCTGGTGG CAGGAGGGGC TCCCCCTGTT
2851 GCGGGTGAGG CGGCTGCTCT CTATTTTCAG ATGTTGCTGT AGAAATAAAG
2901 ACGGTTTAAA TCTGAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

98288268:

Two kinesin light chain genes in mice. Identification and characterization of the encoded proteins.

Peptide information for frame 3

ORF from 144 bp to 2009 bp; peptide length: 622
 Category: strong similarity to known protein
 Prosite motifs: RGD (502-505)
 KINESIN_LIGHT (223-265)
 KINESIN_LIGHT (265-307)

```

1 MAMMVFPEE KLSQDEIVLG TRAVIQGLET LRGEHRALLA PLVAPEAGEA
51 EPGSQERCIL LRRSLEAIEL GLGEAQVILA LSSHLAGVES EKQKLRAQVR
101 RLVOENQWLR EELAGTQQKL QRSEQAVAQL EEKQHLLEFM SQIRKLDEDA
151 SPNEEKGDVP KDTLDDLFNP EDEQSPAPSP GGGDVSGQHG GYEIPARLRT
201 LHNLIQYAS QGRYEVAVPL CKQALDLEK TSGHDHPDVA TMLNILALVY
251 RDQNKYKEAA HLLNDALAIR EKTGLGDHPA VAATLNNLAV LYGKRGKYKE
301 AEPLCKRALE IREKVLGKFH PDVAKQLSNL ALLCQNGQKA EEVEYYRRA
351 LEIYATRLGP ODPNVAKTKN NLASCYLKQG KYQDAETLYK EILTRAHEKE
401 FGSVNGDNKP IWMHAEEREE SKDKRRDSAP YGEYGSWYKA CKVDSPTVNT
451 TLRSIGALYR RQKLEAAHT LEDCASRNRK QGLDPASQTK VVELLKDGSG
501 RRGDRRSSRD MAGGAGPRSE SDLEDVGPTA EWNGDGSGL RRGSGFGKLR
551 DALRRSSEML VKKLQGGTPQ EPPNPRMKRA SSLNFLNKS VEEPTQPGGTG
601 LSDSRTLSSS SMDLSRRSSL VG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4h6, frame 3

TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus kinesin light chain 2 (Klc2) mRNA, complete cds., N = 1, Score = 2824, P = 4e-294

PIR:I53013 kinesin light chain - human, N = 1, Score = 1927, P = 4.5e-199

PIR:C41539 kinesin light chain C - rat, N = 1, Score = 1919, P = 3.2e-198

SWISSPROT:KNLC_RAT KINESIN LIGHT CHAIN (KLC)., N = 1, Score = 1919, P = 3.2e-198

>TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus kinesin light chain 2 (Klc2) mRNA, complete cds.
 Length = 599

HSPs:

Score = 2824 (423.7 bits), Expect = 4.0e-294, P = 4.0e-294
Identities = 558/598 (93%), Positives = 572/598 (95%)

```

Query:      1 MAMMVFPREKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLVAPEAGEAEPGSQERCIL 60
            MA MV PREKLSQDEIVLGTKAVIQGLETLRGEHRALLAPL + EAGEAEPGSQERC+L
Sbjct:      1 MATMVLFPREKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLASHEAGEAEPGSQERCLL 60

Query:     61 LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL 120
            LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL
Sbjct:     61 LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL 120

Query:    121 QRSEQAVAQLEEEKQHLLFMSQIRKLDEASPNEEKGDVPKDLDLDFPNEDEQSPAPSP 180
            QRSEQAVAQLEEEKQHLLFMSQIRKLDE P EEKGDVPKD+LDDLFPNEDEQSPAPSP
Sbjct:    121 QRSEQAVAQLEEEKQHLLFMSQIRKLDE-MLPQEEKGDVPKDSDLDLFPNEDEQSPAPSP 179

Query:    181 GGGDVSGQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA 240
            GGGDV+ QHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA
Sbjct:    180 GGGDVAAQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA 239

Query:    241 TMLNILALVYRDQNKYKEAAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRKYKE 300
            TMLNILALVYRDQNKYK+AAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRKYKE
Sbjct:    240 TMLNILALVYRDQNKYKDAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRKYKE 299

Query:    301 AEPLCKRALEIREKVLGKFHPDVAQQLSNLALLCQNGKAEVEYYYRRALEIYATRLGP 360
            AEPLCKRALEIREKVLGKFHPDVAQQLSNLALLCQNGKAEVEYYYRRALEIYATRLGP
Sbjct:    300 AEPLCKRALEIREKVLGKFHPDVAQQLSNLALLCQNGKAEVEYYYRRALEIYATRLGP 359

Query:    361 DDPNVAKTKNNLASCYLKQGYQDAETLYKEILTRAHEKEFGSVNGDNKPIWMHAEERE 420
            DDPNVAKTKNNLASCYLKQGYQDAETLYKEILTRAHEKEFGSVNG+KPIWMHAEERE
Sbjct:    360 DDPNVAKTKNNLASCYLKQGYQDAETLYKEILTRAHEKEFGSVNGENKPIWMHAEERE 419

Query:    421 SKDKRRDSAPYGEYGSWKACKVDSPTVNTTLRSLGALYRRQGLEAAHTLEDCASTRNRK 480
            SKDKRRD P EYGSWKACKVDSPTVNTTLR+LGALYR +GKLEAAHTLEDCASTR+RK
Sbjct:    420 SKDKRRDRRPM-EYGSWKACKVDSPTVNTTLRSLGALYRPEGKLEAAHTLEDCASTRSRK 478

Query:    481 QGLDPASQTKVVELLKDGSGRRGDRSSRDMAGGAPRSESDLEDVGPTAEWNGDGSGL 540
            QGLDPASQTKVVELLKDGSGR G RR SRD+AG P+SESDLE+ GP AEW+GDGSGSL
Sbjct:    479 QGLDPASQTKVVELLKDGSGR-GHRRGSRDVG---PQSESDLEESGPAAEWSGDGSGL 534

Query:    541 RRGSGFGLKLRDALRRSSEMLVKKLQGGTPQEPNPRMKRASSLNFLNKSVEEPTQPGG 598
            RRGSGFGLKLRDALRRSSEMLV+KLQGG PQEP N RMKRASSLNFLNKSVEEP QPGG
Sbjct:    535 RRGSGFGLKLRDALRRSSEMLVRKLQGGGPQEP-NSRMKRASSLNFLNKSVEEPVQPGG 591

```

Pedant information for DKFZphtes3_4h6, frame 3

Report for DKFZphtes3_4h6.3

```

[LENGTH]      622
[MW]           68934.82
[pI]           6.72
[HOMOL]        TREMBL:AF055666.1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus
kinesin light chain 2 (Klc2) mRNA, complete cds. 0.0
[BLOCKS]       BL00927C Trehalase proteins
[BLOCKS]       BL01160I Kinesin light chain repeat proteins
[BLOCKS]       BL01160H Kinesin light chain repeat proteins
[BLOCKS]       BL01160G Kinesin light chain repeat proteins
[BLOCKS]       BL01160F Kinesin light chain repeat proteins
[BLOCKS]       BL01160E Kinesin light chain repeat proteins
[BLOCKS]       BL01160D Kinesin light chain repeat proteins
[BLOCKS]       BL01160C Kinesin light chain repeat proteins
[BLOCKS]       BL01160B Kinesin light chain repeat proteins
[BLOCKS]       BL01160A Kinesin light chain repeat proteins
[SUPFAM]       tetratricopeptide repeat homology 1e-07
[PROSITE]      RGD 1
[PROSITE]      MYRISTYL 8
[PROSITE]      KINESIN_LIGHT 2
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 5
[PROSITE]      CK2_PHOSPHO_SITE 11
[PROSITE]      TYR_PHOSPHO_SITE 3
[PROSITE]      PKC_PHOSPHO_SITE 7
[PROSITE]      ASN_GLYCOSYLATION 2
[PFAM]         Kinesin light chain repeat
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 12.54 %
[KW]           COILED_COIL 4.98 %

```

Prosites for DKFZphtes3 4h6.3

888

PS00006	568->572	CK2_PHOSPHO_SITE	PDOC00006
PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	610->614	CK2_PHOSPHO_SITE	PDOC00006
PS00007	339->346	TYR_PHOSPHO_SITE	PDOC00007
PS00007	339->347	TYR_PHOSPHO_SITE	PDOC00007
PS00007	424->432	TYR_PHOSPHO_SITE	PDOC00007
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	482->488	MYRISTYL	PDOC00008
PS00008	598->604	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00009	292->296	AMIDATION	PDOC00009
PS00009	499->503	AMIDATION	PDOC00009
PS00016	502->505	RGD	PDOC00016
PS01160	223->265	KINESIN_LIGHT	PDOC00893
PS01160	265->307	KINESIN_LIGHT	PDOC00893

Pfam for DKFZphtes3_4h6.3

HMM_NAME	Kinesin light chain repeat			
HMM	*RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*			
	+ALED+EKT+GHDHPDVATMLN+LALV+R+QNKY+E++ ++N			
Query	223	QALDLEKTSGDHPDVATMLNILALVYRDQNKYKEAAHLLN	264	
50.46	265	306	1	42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain
Alignment to HMM consensus:				
Query		*RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*		
		AL +REKTLG DHP VA LNNLA+++ ++KY+E+E + +		
dkfzphes3	265	DALAIREKTLGKDHPAATLNNLAVLYGKRGKYKEAEPLCK	306	
Query	348	1	42	dkfzphes3_4h6.3 strong similarity to Kinesin light chain
Alignment to HMM consensus:				
HMM	*RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*			
	RALE+REK+LG HPDVA++L+NLAL+C+NQ+K EEVE YY+			
Query	307	RALEIREKVLGKFHPDVAQLSNLALLCQNGKAEVEYYR	348	
39.10	349	390	1	42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain
Alignment to HMM consensus:				
Query		*RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*		
		RALE+ LG D P+VA+ NNLA + Q+KY+++E +Y+		
dkfzphes3	349	RALEIYATRLGPDPPNVAKTKNNLASCYLKQGYQDAETLYK	390	

DKFZphtes3_4o19

group: testes derived

DKFZphtes3_4o19 encodes a novel 1180 amino acid protein with weak similarity to human megakaryocyte stimulating factor and human mucin.

The novel protein contains a cytochrome c family heme-binding site signature.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to megakaryocyte stimulating factor and mucin

complete cDNA, complete cds, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 3767 bp

Poly A stretch at pos. 3757, polyadenylation signal at pos. 3737

```
1  GGCTAGGTTT  AGCTTCAGGG  GCAGCCCAGG  GCAGTGTTC  TGCATATTGC
51  ATGGATGAAA  GGCTGAAGGC  TGCCTCCTCT  TGCAGGCTGG  CTCTGAGAT
101  TGCACCTTCT  TCTCCTGCTA  CTCCTCCAAA  TCTATGACCC  TTCAAGGCAG
151  AGCTGACCTG  TCCGGTAATC  AAGGCAATGC  AGCCGCGCCG  CTAGCTACAG
201  TTCACGAGCC  AGTTGTCACC  CAGTGGGCGG  TGCATCCTCC  AGCCCCCGCT
251  CACCCAGCTC  TCCTGGACAA  AATGGAGAAA  GCGCCTCCAC  AGCCCCAGCA
301  CGAGGGCCTC  AAGTCCAAGG  AGCATCTTCC  GCAACAGCCT  GCCGAAGGCA
351  AGAGGGCGTC  CCGCCGCGTC  CCACGCCTCC  GGGCTGTGGT  CGAGAGCCAG
401  GCCTTCAAGA  ACATCCTGGT  AGACGAGATG  GACATGATGC  ACGCCCGTGC
451  AGCCACGCTC  ATCCAAGCCA  ACTGGAGGGG  CTATTGGCTC  CGGCAGAAGC
501  TGATTTCCCA  GATGATGGCG  GCCAAGGCCA  TCCAGGAGGC  CTGGCGGCGC
551  TTCACAAGA  GACACATCCT  TCACTCCAGC  AAGTCGTTGG  TAAAGAAAAC
601  GAGGGCGGAG  GAGGGGGACA  TACCTTATCA  CGCCCCACAG  CAGGTGCGCT
651  TCCAGCATCC  GGAAGAGAAC  CGCCTTCTGT  CCCC GCCCAT  CATGGTGAAC
701  AAGGAGACCC  AGTTCCCTTC  CTGTGACAAT  CTGGTCTCT  GCAGACCCCA
751  GTCGTCCCCC  CTCCTGCAGC  CCCCAGCAGC  TCAGGGTACC  CCAGAGCCCT
801  GTGTGCAGGG  TCCTCATGCT  GCCAGAGTCC  GGGGGCTGGC  CTTCTTGCCA
851  CACCAGACGG  TCACCATCAG  ATTTCCCTGC  CCAGTGAGTT  TGGACGCAAA
901  ATGCCAGCCA  TGCCTGCTGA  CCAGAACCAT  CAGAAGCACC  TGCCTCGTCC
951  ACATAGAGGG  TGACTCAGTG  AAGACCAAAC  GTGTAAGTGC  CCGGACCAAC
1001  AAAGCCAGGG  CTCGGAGAGC  ACCATTGTCC  AGAAGGTATG  ACCAGGCAGT
1051  TACGAGACCA  TCCAGAGCCC  AAACCCAGGG  CCCTGTGAAA  GCAGAGACCC
1101  CCAAGACCCC  CTTCCAGATA  TGTCCAGGGC  CCATGATCAC  CAAGACTCTA
1151  CTCCAGACAT  ATCCAGTGGT  CTCCTGTGAC  CTGCCACAGA  CATATCCAGC
1201  GTCCACGATG  ACCACCACCC  CACCCAGAC  TAGCCAGATT  CCAAAGTAA
1251  CAATTAATCA  GACCCAGGCC  CAGATGTATC  CGGGGCCAC  AGTGACCAAA
1301  ACTGCACCTC  ACACATGCCC  CATGCCACA  ATGACCAAGA  TCCAGGTACA
1351  CCCCACAGCC  TCCAGAACTG  GCACCCACG  GCAGACATGC  CCTGCGACCA
1401  TCACGGGCAA  GAACCGACCT  CAGGTTTCCC  TTCTGGCTTC  CATCATGAAG
1451  AGCTGCCCCC  AGGTATGCCC  GGGGCCTGCG  ATGGCAAAGA  CCCCACCCCA
1501  GATGCACCCG  GTCACCACCC  CAGCCAAAA  CCCATTGCAA  ACATGTCTGT
1551  CAGCCACAAT  GTCCAAGACT  TCATCCAGA  GGAGCCAGT  TGGGGTGACC
1601  AAGCCTCAC  CCCAGACCCG  CCTGCCAGCC  ATGATAACCA  AGACCCAGC
1651  CCAGTTACGC  TCGGTGGCCA  CCATCCTCAA  GACTCTGTGT  CTGGCTCTC
1701  CAACAGTGGC  AAATGTCAAG  GCTCCACCCC  AAGTGGCGGT  AGCAGCCGGA
1751  ACTCCCAACA  CCTCAGGCTC  CATCCATGAG  AACCCACCCA  AGGCCAAGGC
1801  CACCGTGAAT  GTGAAGCAGG  CTGCAAAGGT  GGTGAAGGCC  TCATCCCCCT
1851  CCTATTGGGC  TGAGGGGAAG  ATCAGGTGCC  TGGCTCAACC  ACATCCGGGA
1901  ACTGGGGTCC  CCAGGGCTGC  AGCTGAGCTT  CCTTTGGAAG  CCGAGAAAA
1951  CAAGACTGGC  ACCAGAAAC  AGGCGAAAC  AGACATGGCA  TTTAAGACCA
2001  GTGTGGCAGT  GGAAATGGCT  GGGGCTCCAT  CCTGGACAAA  AGTTGCTGAG
2051  GAAGGGGACA  AGCCACCTCA  CGTGTATGTG  CCTGTAGACA  TGGCTGTCAC
2101  CCTGCCCCGG  GGACAGCTGG  CTGCCCCACT  GACCAATGCC  TCATCCCAGA
2151  GACATCCACC  CTGCCTGTCC  CAGAGACCAC  TGGCCGCCCC  GCTGACCAAG
2201  GCCTCATCTC  AGGGACATCT  GCCCACTGAG  CTGACCAAGA  CCCCATCCCT
2251  GGCCCATCTG  GACACCTGTC  TGAGCAAGAT  GCATTCCCAG  ACACATCTGG
2301  CCACAGGTGC  CGTGAAGGTC  CAGTCCCAAG  CGCCTCTAGC  CACCTGTCTG
2351  ACCAAGACGC  AGTCCCGGGG  GCAGCCGATC  ACAGACATAA  CCACGTGCCT
2401  CATCCACGCG  CACCAGGCTG  CTGATCTCAG  CAGCAACACC  CACTCCACGG
2451  TGCTCCTAAC  AGGGTCCAAG  GTGTCCAACC  ACGCCTGCCA  CGCCTCGGT
2501  GGCCTCAGCG  CCCCACCTTG  GGCCAAGCCA  GAGGACAGAC  AGACCCAGCC
2551  ACAGCCCCAC  GGACACGTGC  CGGGGAAGAC  CACTCAGGGG  GGACCATGCC
2601  CGGCAGCCTG  TGAGGTCCAG  GGTATGCTGG  TGCCGCCGAT  GGCACCCACC
```

```

2651 GGCCATTCCA CATGCAACGT TGAGTCCTGG GGAGACAACG GAGCCACACG
2701 TGCCCAGCCA TCAATGCCCC GCCAGGCGGT GCCCTGCCAG GAGGACACGG
2751 GCCCCGCGGA CGCTGGTGTG GTTGGTGGCC AATCGTGGAA CCGCGCATGG
2801 GAGCCAGCCA GGGGTGCTGC GTCTGGGAC ACCTGGCGCA ACAAGGCGGT
2851 GGTGCCTCCC AGGCGGTCCG GGGAGCCAAT GGTGTCCATG CAGGCTGCAG
2901 AGGAGATCCG CATCCTCGCA GTGATCACTA TCCAGGCGGG CGTCCGTGGC
2951 TACCTGGCGC GTCGCAGGAT CCGGCTGTGG CACCGGGGGG CCATGGTCAT
3001 CCAAGCTACT TGGCGCGGCT ACCGTGTGCG GCGGAACCTG GCACACCTCT
3051 GCAGAGCCAC CACGACCATC CAGTCTGCCT GCGCGGGCTA CAGCACCCGC
3101 CGGGACCAAG CCCGGCACTG GCAGATGCTC CACCCCGTCA CGTGGGTGGA
3151 GCTGGGCGAG CGGGCCGGGG TCATGTCTGA CCGAAGCTGG TTCCAGGATG
3201 GCAGAGCCAG GACAGTATCT GACCATCGCT GCTTCCAGTC CTGCCAGGCA
3251 CACGCTTGCA GCGTCTGCCA CTCCCTGAGC TCCAGGATCG GGAGCCCGCC
3301 CAGCGTGGTG ATGCTAGTGG GCTCCAGCCC TCGCACCTGT CATACCTGTG
3351 GACGCACACA GCCCACCCTG GTGGTGCAGG GCATGGGCCA GGGCACTGAG
3401 GGGCCCGGGG CAGTGTCTTG GGCCTCCGCC TACCAGCTGG CTGCCCTGAG
3451 TCCCAGGCAG CCGCATCGCC AGGACAAAGC GGCCACAGCC ATCCAGTCCG
3501 CCTGGAGGGG CTTTAAGATC CGCCAGCAGA TGAGGCAGCA GCAATATGGCA
3551 GCGAAGATAG TTCAAGCCAC CTGGCGAGGC CACCATACCC GGAGCTGTCT
3601 GAAGAACACA GAGGCGCTCT TGGGACCAGC AGACCCCTCG GCCAGCTCAC
3651 GGACACATGCA TTGGCCTGGC ATCTAGGACC CTGGCTCCCT GCAGTGGGGA
3701 CTTGCTGGGA GGCACATG GCTCTCTGGG TCTAATGAAT AAAGTCCTCC
3751 ACAGCCTAAA AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3673 bp; peptide length: 1180
 Category: similarity to known protein

```

1  MTLQGRADLS GNQGNAAGRL ATVHEPVVTO WAVHPPAPAH PSLLDKMEKA
51  PPQPOHEGLK SKEHLPPQPA EGKTASRRVP RLRAVVESQA FKNILVDEMD
101 MMHARAATLI QANWRGYWLR QKLISQMAA KAIQEAARRF NKRHILHSSK
151 SLVKKTRAEE GDIPYHAPQ VRFQHFENR LSPPIVMNK ETQFPSCDNL
201 VLCPQSSPL LQPPAAQGT EPCVQGPAA RVRGLAFLPH QTVTIRFPCP
251 VSLDAKQPC LLTRTIRSTC LVHIEGDSVK TKRVSARTNK ARAPETPLSR
301 RYDQAVTRPS RAQTQGPVKA ETPKAPFQIC PGPMITKTL QTYPVVSVTL
351 PQTYPASTMT TTPKTSVPV KVTIITPAQ MYPGPTVTK APHTCPMPTM
401 TKIQVHPTAS RTGTPRQTC ATITAKNRQ VSLLASIMKS LPQVCPGPAM
451 AKTPPQMHPV TTPAKNPLQ CLSATMSKTS SQSPVGVTK PSPQTRLPM
501 ITKTPAQLRS VATILKTLCL ASPTVANVKA PPOVAVAAGT PNTSGSIHEN
551 PPKAKATVNV KQAAKVVKAS SPSYLAEGKI RCLAQPHPGT GVPRAAAELP
601 LEAEKIKTGT QKQAKTDMF KTSVAVEMAG APSWTKVAEE GDKPPHVYVP
651 VDMAVTLPRG QLAAPLTNAS SQRHPPCLSQ RPLAAPLTKA SSQGHLPTEL
701 TKTPSLAHL D TCLSKMHSQT HLTGAVKVQ SQAPLATCLT KTQSRGQPI
751 DITTCILPAH QAADLSSNTH SQVLLTGSKV SNHACQRLGG LSAPPWAKPE
801 DRQTQPQPHG HVPGKTTQGG PCPAACEVQG MLVPPMAPTG HSTCNVESWG
851 DNGATRAQPS MPGQAVPCQE DTGPADAGVV GGQSWNRWE PARGAASWDT
901 WRNKAVVPPR RSGEPMVSMQ AAEEIRILAV ITIQAGVRGY LARRRIRLWH
951 RGAMVIQATW RGYRVRRNLA HLCRATTTIQ SAWRGYSTRR DQARHWQMLH
1001 PVTWVIELGS AGVMSDRSWF QDGRARTVSD HRCFQSCQAH ACSVCHSLSS
1051 RIGSPSPSVVM LVGSSPRTCH TCGRTQPTRV VQGMGQGTG PGAVSWASAY
1101 QLAALSPROP HRQDKAATAI QSAWRGFKIR QQMRQQQMAA KIVQATWRGH
1151 HTRSCCLKNTE ALLGPADPSA SSRHMHWPFI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_4019, frame 2

TREMBL:HSU70136_1 product: "megakaryocyte stimulating factor"; Human
 megakaryocyte stimulating factor mRNA, complete cds., N = 2, Score =
 242, P = 9.6e-16

Query:	281	TKRVSARTNKARAPETPLSRRYDQAVTRPSRAQTQGPVKAE--TPKAPFQICPGPMITKT	338
		TK+ + K AP TP + A TP + P K TP+ P P + T	
Sbjct:	597	TKKPAPTAPKEPAPTTPK----ETAPTTPKKLPTTPEKLAPTTPEKPAPTTPEELAPT	652
Query:	339	LLQTPYVVSVSLPQTYTPASTMTTTPPKTSPV-PKVITIKTFAQMYPGPTVTK-TAPHTCP	396
		+ P TP+ + TP+ +P PK TP+ P PT K TAP T P	
Sbjct:	653	PEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPEPAPTTPEKE--PAPTTKETAP-TTP	709
Query:	397	M---PTMTKIQVHPTASRTGTPRQTCPATITAKNRQVSVLLASIMKSLPQVCPGPAMAKT	453
		PT K+ PT + P++ PT + S + K P GA T	
Sbjct:	710	KGTAPTTLK-EPAPTTPKKPAPKELAPTT----TKEPTSTTSD--KPAPTTPKGTAPT-T	761
Query:	454	PPQMHPVTTPAKNPLQTCLSATMSKTSQRSVPVGWTKPSPQTRLPAMITKTPAQLRSVAT	513
		P+ P TTP K T T T + +P KP+P+ P TK P S	
Sbjct:	762	PKEPAP-TTP-KEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT-TKGPTSTTSDKP	818
Query:	514	ILKTLCLASPTVANVKAPPQVAVAAGTPNTSGSIHENPPKAKATVNV----KQAAKVVKA	569
		T +PT AP A PT E PP +V+ K+ +K+	
Sbjct:	819	APTTPKETAPTTPEPAPTTPKKPA--PTTP----ETPPPTTSEVSTPTTKEPTTIHKS	872
Query:	570	---SSPSYLAEGKIRCLAQPHPGTGVPRAAAELPLEAEKIKTGQKQAKTDMAFKTSVAV	626
		S+P AE + L GVP + P + T T K T+ +T+	

Sbjct: 873 PDESTPELSAEPTPKALENSPKKEGVP--TTKTPAATKPEMTTAKDKTTTERDLRTTPET 930

Query: 627 EMAGAPSWTK-VAEEGDKPPHVYPVDMAVTLPRGQLAAPLTNASSQRHPPCLSQRLPAA 685
A AP TK A +K + +T Q+ + T ++ L LA

Sbjct: 931 TTA-APKMTKETATTTEKT-----TESKITATTQVTSTTTQDTPFKITTLKTTTLAP 983

Query: 686 PLTKASSQGHLPTELTKTPSLAHLDTCLSKMHSQTHLATGAVKVQS-----QAPLATCLT 740
+T + + TE+ P +T K + AT K Q + P +T

Sbjct: 984 KVT-TTKKITTTEIMNKPE----ETAKPKDRATNSKAT-TPKPQKPTKAPKKTSTKKP 1037

Query: 741 KTQSR-GQPITDIT---TCLIPAHQAADLSSNTHSQVLLTGSKVSNHACQRLGGLSAPP 795
KT R +P T T T +P + Q ++ N + S

Sbjct: 1038 KTMPRVKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDA 1097

Query: 796 W-AKPEDRQTQOPHGHVPGKTTQGGPCPAACEVQGMVPPMAPTGHSTCN 845
A+ E +PH +P T P QG+++ PM + CN

Sbjct: 1098 GGAEGETPHMLLRPHVFMPEVTPDMDYLPVFN-QGIINPMLSDETNIEN 1147

Score = 198 (29.7 bits), Expect = 2.3e-11, Sum P(2) = 2.3e-11
Identities = 142/513 (27%), Positives = 200/513 (38%)

Query: 204 RPQSSPLLQPPAAQGTPEPCVQGPAAARVRLAFLPHQVTIIRFPCPVSLDAKQPCLLT 263
R + P +PP G + H V+ + P L

Sbjct: 207 RTKKKPTPKPPVDEAGSGLDNGDFKVTTPDTSTTQHNVSTSPKITTAKPINPRPSLPP 266

Query: 264 R--TIRSTCLVHIEGDSVKTNRVSARTNKARAP---ETPLSRRYDQAVTRPSR---AQTO 315
T + T L + +V+TK + TNK + E S + Q++ + S A T

Sbjct: 267 NSDTSKETSILTVNKETTETKETT-TNKQSTDGKEKTSKAKETQSIKTSKADLAPTS 325

Query: 316 GPVKAETPKAPFQICPGPMITKTLQTYPVVSVTLPTQTPASTMTTTPPKTSPVPKVITII 375
+ TPKA GP +T T + P T P+ PAST TP + +P +

Sbjct: 326 KVLAKPTPKAE-TTKGPALT-TPKEPTP---TTPKE-PAST---TPKEPTTTIKSAP 375

Query: 376 KTPAQMYPGPTVTKTAPHTC--PMPTMTKIQVHTASRTGTPTQTC-PATITAKNRQVQS 432
TP + P PT TK+AP T P PT TK + PT + P T PA T K+ P

Sbjct: 376 TTPKE--PAPTTKSAPTTPKEPAPTTTK-EPAPTTPKEPAPTTTKAPAPTTTKSAPTTP 432

Query: 433 ---LLASIMKSLPQVCPGPMAPKTPPMHPVTTPAKNPLQTCLSATMSKTSQSRSPVGV 489
+ K P PA TP + P TTP K P T + T + +P

Sbjct: 433 KEPAPTTPKKPAPTTPKEPAPT-TPKEPTP-TTP-KEPAPTTKEPAPT-TPKEPAPTAPK 488

Query: 490 KPSPQT-RLPAMIT-KTPAQLRSVA---TILK---TLCLASPTVANVKAPQVAVAGT 540
TP + P T + PA T K PA + T K T ++PT AP A T

Sbjct: 489 KPAPTTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPT 548

Query: 541 PNT-SGSIHENP---PKAKATVNVKQAAKV--KASSPSYLAEGKIRCLAQPHPGTGVPR 594
P S + + P PK A K+ A K +P+ E +P P P+

Sbjct: 549 PKEPSPTTTKEPAPTTTPKEPAPTTPKKPAPTTPKEPAPTTTPKEPAPTTTKKPAPTA--PK 606

Query: 595 AAAELPLEAEIKTGTQKQAKTDMAFKTSVAVEMAGAPSWTK-VAEEGDKPPHVYPVDM 653
A P ++ T K+ K + AP+ +A + P P +

Sbjct: 607 EPA--PTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPPELAPTTPEEPTPTTPEEP 664

Query: 654 AVTLPRGQLAAPLTNASSQRHPPCLSQRLPAAPLTKASSQGHLPTELTKTPSLAHLDT 712
A T P+ AAP T + P P + P AP T P E T T

Sbjct: 665 APTTPKA--AAPNT----PKEPAPTTPKEP--APTTPKEPAPTTPKETAPTTPKGTAPT 716

Query: 713 LSK 715
L +

Sbjct: 717 LKE 719

Score = 108 (16.2 bits), Expect = 4.3e-02, Sum P(2) = 4.3e-02
Identities = 60/214 (28%), Positives = 85/214 (39%)

Query: 265 TIRSTCLVHIEGDSVKTNRVSAR-TNKA--RAPETP-LSRRYDQAVTRPSRAQTQGPVKA 320
T + +H D T +SA T KA +P+ P + A T+P T

Sbjct: 862 TTKEPTTIHKSPDE-STPELSAEPTPKALENSPKKEGVPVTKTPAATKPEMTTAKDKTT 920

Query: 321 ETP--KAPFQICPGPMITK-TLLQTYPVVSVTLPTQTPASTMTTTPPKTSPVPKVITIIKT 377
E P P +TK T T + T T TTT T+P K+T +KT

Sbjct: 921 EROLRTTPETTTAAPKMTKETATTTEKTTESKITATTQVTSTTTQD-TTPF-KITTLKT 978

Query: 378 PAQMYPGPTVTK---TAPHTCPMPTMT-KIQVHTASRTGTPTQTCPATITAKNRQVSL 433
+ P T TK T P T K + T S+ TP+ P A +P +

Sbjct: 979 TT-LAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPQKPTK--APKKPTSTK 1035

Query: 434 LASIMKSL--PQVCPGPA-MAKTPPMHPVTTPAKNPLQT 470
M + P+ P P M T P+++P + A+ LQT

Sbjct: 1036 KPMTMPRVKPKTTPTPRKMTSTMPELNPTSRIAEAMLQ 1075

Score = 56 (8.4 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12

Pedant information for DKFZphtes3_4o19, frame 2

Report for DKFZphtes3_4o19.2

894

SEQ MTLQGRADLSGNQNAAGRLATVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLK
SEG
PRD ccccccecccccccccecc

SEQ SKEHLPQQAEGKTASRRVPRLRVAVESQAFKNILVDEMDMMHARAATLIQANWRGYWLR
SEG
PRD cccccccccccccccccchhhhhhhhhhhhhheehhhhhhhhhhhhhhhccchhhh

SEQ QKLISQMMAAKAIQEAWRRFNKRHILHSSKSLVKKTRAEGDIPYHAPQQVRFQHPENR
SEG
PRD hhhhhhhhhhhhhhhhhhhhhheeeccccchhhhhhhcccccccccccccccccccc

SEQ LLSPPIMVNETQFPSCDNLVLCRPOSSPLLQPPAAQGTPEPCVQGPAAARVRGLAFLPH
SEG
PRD ecccccecc

SEQ QTVTIRFPVPSVSLDAKQPCLLTRTIRSTCLVHIEGDSVTKRVSARTNKARAPETPLSR
SEG
PRD eeeeecc

SEQ RYDQAVTRPSRAQTQGPVKAETPKAPFQICPGPMITKTLQTYPVVSVTLPTQYPASTMT
SEG
PRD ccccccecc

SEQ TTPPKTSPVQKVTIITKPAQMPGPTVTKTAPHTCPMPTMTKIQVHPTASRTGTPTQCP
SEG
PRD ccc

SEQ ATITAKNRQVSLASIMKSLPQVCPGPAMAKTTPQMHPVTTPAKNPLQTCLSATMSKTS
SEG
PRD ccc

SEQ SQRSPVGVTKPSQTRLPAKITKPAQLRSVATILKTLCLASPTVANVKAPQVAVAAAGT
SEG
PRD ccc

SEQ PNTSGSIHENPPKAKATVNVKQAAKVVKASSPSYLAEGKIRCLAQHPGTGVPRAAAELP
SEG
PRD ccc

SEQ LEAEKIKTGTQKQAKTDMAFKTSVAVEMAGAPSWTKVAEEGDKPPHYVVPVDMAVTLPRG
SEG
PRD ccc

SEQ QLAAPLTNASSQRHPCLSRPLAAPLTKASSQGHLPTELTKTPSLAHLDTCLSKMHSQT
SEG
PRD ccc

SEQ HLATGAVKVSQAPLATCLTKTQSRGQPITDITCLIPAHQAADLSSNTHSQVLLTGSKV
SEG
PRD ccccccecc

SEQ SNHACQRLGGLSAPPWAKPEDRQTQPPHGHVFGKTTQGGPCPAACEVQGMVPPMAPTG
SEG
PRD ccc

SEQ HSTCNVESWGDNGATRAQPSMPGQAVPCQEDTGPDAGVVGGSWNRAWEARGAASWDT
SEG
PRD ccc

SEQ WRNKAVVPFRSGEPMVSMQAAEIRILAVITIQAGVRGYLARRRIRLWHRGAMVIQATW
SEG
PRD cccccccccccccccccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhh

SEQ RGYRVRNLAHLCRATTTIQSAWRGYSTRDQARHWQMLHPVTWVELGSRAGVMSDRSWF
SEG
PRD hheeeccccchhhhhhhhh

SEQ QDGRARTVSDHRCFQSCQAHACSVCHSLSSRIGSPPSVVMVVGSSPRTCHTCGRTOPTRV
SEG
PRD hcccccecc

SEQ VQGMGQTEGPGAVSWASAYQLAALSPPRPHRQDKAATAIQSAWRGFKIRQOMRQQQMAA
SEG
PRD eccccccccccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ KIVQATWRGHHTRSCLKNTEALLGPADPSASSRHHMWPGI
SEG
PRD hhhhhhhccccchhhhhhhhhcccccccccccccccccccc

Prosites for DKFZphtes3_4ol9.2

PS00001	542->546	ASN_GLYCOSYLATION	PDOC00001
PS00001	668->672	ASN_GLYCOSYLATION	PDOC00001
PS00004	282->286	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	244->247	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	278->281	PKC_PHOSPHO_SITE	PDOC00005
PS00005	281->284	PKC_PHOSPHO_SITE	PDOC00005
PS00005	285->288	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	424->427	PKC_PHOSPHO_SITE	PDOC00005
PS00005	481->484	PKC_PHOSPHO_SITE	PDOC00005
PS00005	610->613	PKC_PHOSPHO_SITE	PDOC00005
PS00005	671->674	PKC_PHOSPHO_SITE	PDOC00005
PS00005	679->682	PKC_PHOSPHO_SITE	PDOC00005
PS00005	900->903	PKC_PHOSPHO_SITE	PDOC00005
PS00005	959->962	PKC_PHOSPHO_SITE	PDOC00005
PS00005	987->990	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1015->1018	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1049->1052	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1065->1068	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1106->1109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1146->1149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1171->1174	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	546->550	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	988->992	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1003->1007	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1027->1031	CK2_PHOSPHO_SITE	PDOC00006
PS00008	11->17	MYRISTYL	PDOC00008
PS00008	14->20	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	591->597	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	777->783	MYRISTYL	PDOC00008
PS00008	853->859	MYRISTYL	PDOC00008
PS00008	878->884	MYRISTYL	PDOC00008
PS00008	882->888	MYRISTYL	PDOC00008
PS00008	1008->1014	MYRISTYL	PDOC00008
PS00008	1053->1059	MYRISTYL	PDOC00008
PS00008	1083->1089	MYRISTYL	PDOC00008
PS00190	1042->1048	CYTOCHROME_C	PDOC00169

(No Pfam data available for DKFZphtes3_4ol9.2)

DKFZphtes3_50j4

group: testes derived

DKFZphtes3_50j4 encodes a novel 187 amino acid protein proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin rich protein

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1186 bp

Poly A stretch at pos. 1176, polyadenylation signal at pos. 1126

```

1  CACTGGGCGT CTGAAGCTCA GAGCTCACCC CTGAGATGGG CTCTCCTAGG
51 CCTCCTGGGA TGAGGGAGCC ACCAGGACCC AGTGCTGTGA TGCCTGCTCT
101 TCCCTCTACC AGCACCTGCC CGCCCAGAGA CCAGGGCACC CCTGAAGTCC
151 AGCCCACCCC TGCAAAGGAC ACATGGAAGG GCAAGCGGCC TCGATCCCAG
201 CAGGAGAACC CAGAGAGCCA GCCTCAGAAG AGGCCACGCC CCTCAGCCAA
251 GCCCTCCGTC GTAGCTGAGG TCAAGGGCAG CGTCTCGGCC AGCGAACAGG
301 GCACCTTGAA TCCCACGGCT CAAGACCCCT TCCAGCTCTC CGCTCCTGGC
351 GTCTCCTTGA AGGAGGCTGC AAATGTTGTG GTCAAGTGCC TCACCCCTTT
401 CTACAAGGAG GGCAAGTTTG CTTCCAAGGA GTTGTTTAAA GGCTTTGCCC
451 GCCACCTCTC AACTTGTCTG ACTCAGAAGA CCTCTCCTGG AAGGAGCGTG
501 AAAGAAGAGG CCCAGAACCT CATCAGGCAC TTCTTCCATG GCCGGGCCCG
551 GTGCGAGAGC GAAGCTGACT GGCATGGCCT GTGTGGCCCC CAGAGATGAC
601 CAACTGTCTG CTGGGCAGGG CCCGCGTCCT CCCCAGATT CTAGCATGGG
651 TCATCCTGGG CCTCACCTGC TGATGCCAGG GCCATCGTCT TTTCTCAGTC
701 CTTCTCCTTT CCAACCATAC TTGGCTTTGG GGATGACCCC AGACACCCCC
751 TGAATCCAGG TCAGAGGTCA GCCCACCTTT CTTTCTGCTT GCAAAGCCTA
801 TAGACCCCTC TCAGAGCGGT CCTCATGGCT GGGTTTCTG GGACACATGT
851 CGAGGACAGA AGGTGGAGGG TGGTGGAGCT GCTGCTGGAA GAAGGGGAAG
901 GAAGAGTGGC CCCTCCCGGA GTTCTAAGTC AGGATGAGGC CCACCTGTCC
951 AAGGTATCGG AACCTACCCA GGGGACCCTC AGATCCTCCA CCCACTCCCC
1001 CATCCATTAC GATGCCAGCT TCCAGCCTTG CCCAGGTCAG AGCTGTGGCA
1051 GAGGAGAGGC AGCCAGGCCC TGTTCTGCTC CAGCTCCTGC TCAGGAAGGC
1101 CAGGCCTGAC AGATGTTTGG GAGAGGAATA AAGTTGTGTT GTTGTGGGGC
1151 ATGCAGGCGT GCACACAGCC CTTTCAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 36 bp to 596 bp; peptide length: 187
 Category: putative protein

```

1  MGSPPRPQMR EPPGPSAVMP ALPSTSTCPP RDQGTPEVQP TPAKDTWKKG
51  RPRSQQENPE SQPQKRPRPS AKPSVVAEVK GSVSASEQGT LNPTAQDPFQ
101 LSAPGVSLKE AANVVVKCLT PFYKEGKFAS KELFRGFARH LSHLLTQKTS
151 PGRSVKEEAQ NLIRHFFHGR ARCESEADWH GLCGPQR

```

BLASTP hits

Entry MMU92455_1 from database TREMBL:

product: "WW domain binding protein 7"; Mus musculus WW domain binding protein 7 mRNA, partial cds.
Score = 134, P = 6.9e-08, identities = 45/125, positives = 56/125

Alert BLASTP hits for DKFZphtes3_50j4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_50j4, frame 3

Report for DKFZphtes3_50j4.3

```

[LENGTH]      187
[MW]           20353.06
[pI]           9.76
[PROSITE]      MYRISTYL      1
[PROSITE]      AMIDATION     1
[PROSITE]      CK2_PHOSPHO_SITE      6
[PROSITE]      PKC_PHOSPHO_SITE      6
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      8.56 %

SEQ      MGSPPRPPGMREPPGPSAVMPALPSTSTCPPRDQGTPEVQPTPAKDTWKGKRPSSQENPE
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      SQPQKRPRPSAKPSVVAEVKGSVSASEQGTINPTAQDPFQLSAPGVSLKEAANVVVKCLT
SEG      .....
PRD      cccccccccccccchhhhhccccccccccccccccccccccccccccccccchhhhhhheeecc

SEQ      PFYKEGKFASKELFKGFARHLSHLLTQKTSPGRSVKEEAQNLI RHFFHGRARCESEADWH
SEG      .....
PRD      cccccccchhhhhhhhhhhhhhhhhheeeccccchhhhhhhhhhhhhhhccchhhhhhhhh

SEQ      GLCGPQR
SEG      .....
PRD      ccccccc

```

Prosite for DKFZphtes3_50j4.3

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	70->73	PKC_PHOSPHO_SITE	PDOC00005
PS00005	107->110	PKC_PHOSPHO_SITE	PDOC00005
PS00005	146->149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81->87	MYRISTYL	PDOC00008
PS00009	48->52	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_50j4.3)

DKFZphtes3_50n06

group: testes derived

DKFZphtes3_50n06 encodes a novel 186 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1095 bp

Poly A stretch at pos. 1085, polyadenylation signal at pos. 1061

```
1 CAAGACCTC GGAGCCAAGA AACAACTG AGTTCCAGAT TTCGGAAGGT
51 TCACGAGTGT TGCCGACACG CCTCCCAAC TGCAGACATC CTCCTGGAG
101 GACCTGCTGT GCTCACATGC CCCCTGTCC AGCGAGGACG ACACCTCCCC
151 GGGCTGTGCA GCGCCCTCCC AGGCACCTT CAAGGCCTTC CTCAGTCCCC
201 CAGAGCCACA TAGCCACCGA GGCACCGACA GGAAGCTGTC CCCGCTCCTG
251 AGCCCTTGC AAGACTCACT GGTGGACAAG ACCCTGCTGG AGCCAGGGA
301 GATGGTCCGG CCTAAGAAGG TGTGTTTCTC GGAGAGCAGC CTGCCACCG
351 GGGACAGGAC CAGGAGGAGC TACTACCTCA ATGAGATCCA GAGCTTCGCG
401 GCGCCCGAGA AGGACGCGCG CGTGGTGGGC GAGATCGCCT TCCAGCTGGA
451 CCGCCGCATC CTGGCCTACG TGTCCCGGG CGTGACGCG CTCTACGGCT
501 TCACGGTGGC CAACATCCCC GAGAAGATCG AGCAGACCTC CACCAAGTCT
551 CTGGACGGCT CCGTGGACGA GAGGAAGCTG CGCGAGCTGA CGCAGCGCTA
601 CTTGGCCCTG AGCGCGCGCC TGGAGAAGCT GGGCTACAGC CGCGACGTGC
651 ACCCGCGGTT CAGCGAGTTC CTCATCAACA CCTACGGAAT CCTGAAGCAG
701 CGGCCCGACC TGCGCGCCAA CCCCTGCAC AGCAGCCCG CCGCGCTGCG
751 CAAGCTGGTC ATCGACGTGG TGCCCCCAA GTTCCTGGGC GACTCGCTGC
801 TGCTGCTCAA CTGCCTGTGC GAGCTCTCCA AGGAGGACGG CAAGCCCTC
851 TTGCGCTGGT GAGCGGCCCC GCGCCGCGC CCTTGCTGCG AGTAAACGCG
901 TTTGTTTCAA CCGGGGCGCG CGGTGCCTCC TGCGCGTCCC CCGGAGGGG
951 AAAGGGCGCG GTCCCCGCG CGCGAGGCCA GAGAAGGCC CGCTCCACC
1001 GGTGCTGGGC CCGACCGCA GCCCGCGCT GCCCGACCT GCGGAGTGCT
1051 TCTACCCCT CATTAAAATC ATCCGTTTGC TTGTCAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 302 bp to 859 bp; peptide length: 186

Category: putative protein

Classification: no clue

```
1 MVRPKKVCFS ESSLPTGDRT RRSYYLNEIQ SFAGAEDKAR VVGEIAFQLD
51 RRILAYVFPQ VTRLYGFTVA NIPERIEQTS TKSLDGSVDE RKLRELTQRY
101 LALSARLEKL GYSRDVHPAF SEFLINTYGI LKQRPDLRAN PLHSSPAALR
151 KLVIDVVPFK FLGDSLLLLN CLCELSKEDG KPLFAW
```

BLASTP hits

No BLASTP hits available

No Alert BLASTP hits found

Pedant information for DKFZphtes3_50n06, frame 2

Report for DKFZphtes3_50n06.2

```
[LENGTH]      186
[MW]           21049.39
[pI]           9.28
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      5.38 %
```

```
SEQ      MVRPKKVCFSSESLPTGDRTRRSYYLNEIQSFAGAEKDARVVGEIAFQLDRRILAYVFPG
SEG      .....
PRD      cccccceccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
```

```

SEQ      VTRLYGFTVANI PEKIEQTSTKSLDGSVDERKLRRLTQRYLALSARLEKLGYSRDVHPAF
SEG      .....
PRD      ceeeeeeeeeeccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccch

```

```

SEQ      SEFLINTYGILKQRPDLRANPLHSSPAALRKLVIDVVPKFLGDSL LLLNCLCELSKEDG
SEG      .....XXXXXXXXXX.....
PRD      hhhhhhhcceeccccccccccccccccchhhhhhhhhccccccccchhhhhhhhhhhcccc

```

```

SEQ      KPLFAW
SEG      . . . . .
PRD      CCCCCC

```

(No Prosite data available for DKF2phtes3_50n06.2)

(No Pfam data available for DKFZphtes3_50n06.2)

DKFZphtes3_50n23

group: testes derived

DKFZphtes3_50n23 encodes a novel 499 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits
(from other testis librarys) testis specific cDNA?

Sequenced by DKFZ

Locus: unknown

Insert length: 1907 bp
Poly A stretch at pos. 1897, polyadenylation signal at pos. 1872

```
1 GGGCACCAGC CACTTCCAC CATGACTGTG CGCTCGAGGG TCGCAGATGT
51 GTTCGGCAGC AAGGACACTG AGAGCCTTGA GCCTGTGCTT TTACCCTTAG
101 TAGATCGCAG GTTTCCTAAG AAATGGGAAA GACCGGTGGC AGAAAGCTTA
151 GGCCACAAAG ACAAGACCA GGAGGACTAC TTCCAGAAAG GAGGACTCCA
201 AATTAAGTTC CACTGTAGCA AGCAGCTGTC TCTAGAGAGC TCCAGGCAGG
251 TGACCTCTGA GAGCCAAGAG GAGCCCTGGG AGGAGGAATT CGGCCGGGAG
301 ATGCGGAGGC AGCTGTGGCT GGAGGAGGAG GAGATGTGGC AGCAGCGGCA
351 GAAGAAGTGG GCCCTGCTGG AGCAGGAGCA TCAGGAGAAG CTGCGGCAGT
401 GGAATCTGGA AGACCTGGCC AGGGAGCAAC AGCGGAGATG GGTCCAGCTA
451 GAAAAGGAGC AGGAGAGCCC ACGGAGAGAG CCAGAGCAGC TAGGGGAGGA
501 TGTGGAGAGG AGGATCTTCA CACCCACCAG TCGATGGAGG GACTTGGAGA
551 AGGCAGAGCT ATCATTAGTG CCTGCCCCAA GCGGACCCA ATCTGCTCAC
601 CAAAGCAGGA GGCCACACTT GCCCATGTCT CCTAGTACCC AGCAGCCTGC
651 CCTGGGAAAG CAGAGACCTA TGAGTTCAGT GGAGTTTACC TACAGACCAC
701 GGACCCGCCG AGTTCCCA CAAGCCCAAGA AATCTGCCTC CTTTCTGTCT
751 ACTGGGACAT CCATCCGAAG GCTGACCTGG CCCTCTTTGC AGATATCCCC
801 TGCAAAATATT AAGAAGAAGG TGTACCACAT GGACATGGAG GCCCAGAGGA
851 AGAACCTGCA GCTCCTGAGT GAGGAGTCTG AGTTGAGGCT GCCCCACTAC
901 CTGCGCAGCA AAGCACTGGA GCTCACCACC ACCACCATGG AGCTGGGCGC
951 GCTCAGGCTG CAGTACCTGT GCCATAAGTA CATCTTCTAT AGACGCCTCC
1001 AGAGCCTCCG GCAAGAAGCG ATCAACCATG TACAAATCAT GAAAGAAACG
1051 GAGGCTTCCT ACAAGGCCCA GAACCTCTAC ATCTTCTTGG AAAACATTGA
1101 CCGCCTGCAG AGTCTCAGGC TGCAGGCCTG GACGGACAAG CAGAAGGGGC
1151 TGGAGGAGAA GCACCGAGAG TGCCTGAGCA GCATGGTGAC CATGTTCCCC
1201 AAGCTCCAGC TGGAGTGGAA CGTTCACCTG AACATCCCTG AGGTACCTTC
1251 GCCAAAGCCA AAGAAATGCA AGTTGCCTGC AGCCTCACCC CGGCACATCC
1301 GCCCCAGTGG CCCCACCTAC AAGCAGCCCT TTCTGTCTAG GCACCGGGCA
1351 TGTGTGCCCC TGCAGATGGC CCGCCAACAG GGGAAAGCAG TGGAGGCTGT
1401 CTGGAAGACC GAGGTGGCCT CCTCCAGTTA CGCAATAGAA AAAAGACCC
1451 CTGCCAGCCT TCCCCGGGAC CAGCTGAGGG GACACCCAGA TATTCCCCGG
1501 CTGTTGACAC TGGACGTGTA GTCCTCCTGC CACAAAAGCC TGAACCTCCT
1551 GAAGGCCCCG TAAGCGCCTC AGCGAACCAG AGGAAGGAAT GCCAGGAACC
1601 TACAAATGAA TCCGCTTAGC TTGTTCAAAA AAAGTCAAGC GAGTCACTCC
1651 CTGGAACCCA AATAAGCCAG AAGGATCAAG ACAGCCCCAG TCTCCACTGC
1701 ATCCCTCAGC CAGTGATTCT CAACCTTCTG AGGGACGGAA ACCCAGAGAG
1751 AACTTGGTCA AAATGCAGGT TCCCAGCTGG TGCTTTTAAA GAAACCTCTC
1801 GGGGGTTGCT GAGTACTCCT AGAACTTTGA GAAACACTGC TTCCCTCCTG
1851 CAGTCCCCAA ACTCTACATT TTAATAAAAT AGAGGTTGGT TTATTTTAAA
1901 AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 22 bp to 1518 bp: peptide length: 499
 Category: similarity to known protein
 Classification: no clue

```

1  MTVRSRVADV FGSKDTESLE FVLLPLVDRR FPKKWERPVA ESLGHKDKDQ
51 EDYFQKGGGLQ IKFHCSKQLS LESSRQVTSE SQEEPWEEEF GREMRRLWL
101 EEEEMWQQRQ KKWALLEQEH QEKLRQWNLE DLAREQQRWV VQLEKEQESP
151 RREPEQLGED VERRIFTPTS RWRDLEKAE SLVPAPSRTO SAHQSRPHL
201 PMSPTQQA LGKQRPMSV EFTYRPTRR VPTKPKKSAS FVPTGTSIRR
251 LTWPSLQISP ANIKKKVYHM DMEAQRKNLQ LLSESELRL PHYLRSKALE
301 LTTTTMELGA LRLQYLCHKY IFYRRLQSLR QEAINHVQIM KETEASYKAQ
351 NLYIFLENID RLQSLRLQAW TDKQKGLEEK HRECLSSMVT MFPKLQLEWN
401 VHLNIPEVTS PKPKCKLPA ASPRHIRPSG PTYKQPFSLR HRACVPLQMA
451 RQQGQKMEAV WKTEVASSY AIEKKT PASL PRDQLRGHPD IPRLLTLDV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_50n23, frame 1

PIR:S28589 trichohyalin - rabbit, N = 1, Score = 134, P = 5.3e-05

TREMBLNEW:AF132479_1 product: "Ese2L protein": Mus musculus Ese2L
 protein mRNA, complete cds., N = 1, Score = 130, P = 0.00017

>PIR:S28589 trichohyalin - rabbit
 Length = 1,407

HSPs:

Score = 134 (20.1 bits), Expect = 5.3e-05, P = 5.3e-05
 Identities = 88/354 (24%), Positives = 154/354 (43%)

```

Query: 29 RRFPPKWERPVAESLGHKDKDQEDYFQKGGGLQIK-FHCSKQLSLESSRQVTSESQEEPWE 87
      R++ K +R + L + ++E ++ G + F +QL +++ E +EE +
Sbjct: 165 RQYRDKEQLRQLEERRAEQQRLRRRGRDAEEFIEEQRLRRREQQLKRELREEEQ 224

Query: 88 EEFGREMRRLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRWVQLEKEQ 147
      RE + L+EEE RQ++W E Q++LR+ LE++ RE+++R Q E+ +
Sbjct: 225 RREERREQHERA-LQEEEEQLRQRRWRE-EPREQQQLRR-ELEEI-REREQRLEQEERRE 280

Query: 148 ESPRPEQLGEDVERRIFTPTSRWRDLEKAE SLVPAPSRTO SAHQSRPHL PMSPTQ 207
      + RRE ++L E ERR ++ + E L R Q Q R + +
Sbjct: 281 QQLRRE-QRL-EQEERREQLRRELEEI REREQRLEQEERREQRLEQEERREQLKRELE 338

Query: 208 QPALGKQRPMSVVEFTYRPTRRVPTKPKKSASFVPTGTSIRRLTWPSLQISPANIKK-K 266
      + +QR +E R R + + + ++ A G S+ R W S A ++ K
Sbjct: 339 EIREREQR----LEQEER-REQLLAEEVREQAR--ERGESLTR-RWQRQLESEAGARQSK 390

Query: 267 VYHMDMEAQRKNLQLLSESELRLPHYLRSKALELTTTTM-----ELGALRLQYLCHKY 320
      VY +R+ Q L ++ E R R + LE E R Q L +
Sbjct: 391 VYS---RPRRQEEQLRQDQERR-QRQERERELEEQARRQQWQAEESERRRQRLSARP 446

Query: 321 IFYRRLQSLRQEAINHVQIMKETEASYKAONLYI-FLENIDRLQSL-RLQAWTDKQKGLE 378
      R Q +E Q +E E ++ + FLE ++LQ R Q ++ E
Sbjct: 447 SLRER-QLRAEERQEQRFRREEEQRRRRQELQFLEEEQLQRRERAQQLQEEDSFQE 505

Query: 379 EKHR 382
      ++ R
Sbjct: 506 DRER 509

```

Score = 119 (17.9 bits), Expect = 2.2e-03, P = 2.2e-03
 Identities = 79/357 (22%), Positives = 150/357 (42%)

```

Query: 33 KKWERPVAESLGHKDKDQEDYFQKGGGLQIKFHCSKQLSLESSRQVTSESQEEPWEEFGR 92
      ++ E+ + + K +++E Q+ ++ +Q R+ + + + EE+F +
Sbjct: 990 RREEQLRQERDRKFREEEQQLQE---REEERLRRQERDRKFREEERQLRQLEEQFRQ 1046

Query: 93 EMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRWVQLEKEQESPRR 152
      E R+ LEE+ + Q++++K L QE K R+ E+ R +Q R QL +E++ R
Sbjct: 1047 ERDRKFRLEEQ-IRQKEEK-QLRRQERDRKFRE---EEQRRRQEREQQLRRERDRKFR 1101

Query: 153 EPEQLGEDVERRIFTPTSRWRDLEKAE SLVPAPSRTO SAHQSR--RPHLPMSPTQQA 210

```

Sbjct: 1102 E EQL ++ E R R L + E L + + + R R + +++
 EEEQLQREERLRQERARKLREEE-QLLRREEQLLRQERDRKFREEEQLLQSESEER 1160

Query: 211 LGKQ---RPMSSVEFTYRPRTRVPKPKKSASFVGTGTSIRRLTWPSLQISPANIKKKV 267
 L + Q R + E + R + +++ +R+ Q ++++

Sbjct: 1161 LRRQERERKLREEEQLLQREERLRQERARKLREEEQLLRQEEQLRQERARKLREEE 1220

Query: 268 YHMDMEAQ-----RKNLQLLS-ESELRPHYLRSKALELTTTMMELGALRLQYL 316
 + E Q + L E R+ QLL EE ELR + + E E LR Q

Sbjct: 1221 QLLRQEEQLRQERDRKFREEEQLLRREEQLRRERDRKFREEEQLLQREERLRQER 1280

Query: 317 CHKYIFYRRLQSLRQEAINHVMKETEASYKAQONLYIFLENIDRLQ-SLRQAWTDKQK 375
 K + L E ++ +E + Y+A+ + E RL+ LR + +++

Sbjct: 1281 ARK--LREEEQLLFEEQEEQLRQERDRRYRAEEQFAREEKSRLERELRQEEEQRRR 1338

Query: 376 GLEEKHRE 383
 E K RE

Sbjct: 1339 ERERKFRE 1346

Score = 109 (16.4 bits), Expect = 1.9e-01, P = 1.7e-01
 Identities = 37/113 (32%), Positives = 60/113 (53%)

Query: 67 KQLSLESSRQVTSSESQ--EEPWEEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK 124
 +QL E R+ E Q +E EE R+ R + EEE++ Q+R+++ L QE + KL

Sbjct: 764 QLLRRERDRKFREEEQLLQREERLRQERERKLREEEQLLQREERLRLRQERERKL 822

Query: 125 RQWNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
 R+ E L +E++ ++ +E+E RE EQL E+ + R R L + E

Sbjct: 823 REE--EQLLQREERLR-RQERERKLREEEQLLRQEEQL--RQERARKLREEE 872

Score = 107 (16.1 bits), Expect = 3.0e-01, P = 2.6e-01
 Identities = 35/109 (32%), Positives = 61/109 (55%)

Query: 71 LESSRQVTSSESQEEPWE-EEFGREMRRQL---WLEEEEMWQQRQKKWALLEQEHQEK 126
 L Q+ ES+EE +E +++RR+ + EEE++ Q+R+++ L QE + KLR+

Sbjct: 742 LREEEQLLQSEERLRQEREQQLRRERDRKFREEEQLLQREERLRLRQERERKLRE 800

Query: 127 WNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
 E L +E++ ++ +E+E RE EQL ++ E R R L + E

Sbjct: 801 E--EQLLQREERLR-RQERERKLREEEQLLQREERLRQERERKLREEE 850

Score = 104 (15.6 bits), Expect = 9.4e-02, P = 9.0e-02
 Identities = 84/339 (24%), Positives = 149/339 (43%)

Query: 67 KQLSLESSRQVTSSESQEEPWEEEFGRMRRQL-WLEEEEMWQQRQKKWALLEQ--HQEK 123
 +QL E ++ +EE EE RE R++L +LEEEE Q+R++ L E++ ++

Sbjct: 451 QLRAEERQEQRQFREE--EEQRERRRQELQFLEEEQLQRRRAQQLEEDSFQEDR 507

Query: 124 LRQWNLEDLAREQRRWVQLEKEQESPRR---EP---EQLGEDVE-RRIFTPTSRWRDL 175
 R+ ++ Q RW QL++E + R +P EQL E+ E +R R R+

Sbjct: 508 ERRRRQEQRPQGTWRW-QLQEEAQRRTLYAKPGQQEQLREEEELQREKRRQEREREY 566

Query: 176 EKAELSLVPAPSRTQSAHQSRPHLPSPSTQPPALGKQPMSSVEFTYRPT---RRV 231
 + E L + + + R + + Q+ L + R + E + R RR

Sbjct: 567 REEE-KLQREDEKRRRQERERQYRELELRQEEQL-RDRKLREEEQLLQREERLRQ 624

Query: 232 PTKPK---KSASFVGTGTSIRRLTWPSLQISPANIKKKVYHMDMEAQRK---NLQLLSEE 285
 + K + +R+ L+ +++++ + E +RK QLL E

Sbjct: 625 ERERKLREEEQLLRQEEQLRQERERKLREEEQLLRREEQLRQERERKLREEEQLLQER 684

Query: 286 SELRLPHYLRSKALE-----LTTTMMELGALRLQYLCHKYIFYRRL-QSLRQEAINHVM-- 337
 E RL R++ L L EL R + L + RR Q LRQE +

Sbjct: 685 EERLRQERARKLREEEQLLRQEEQLRQERERKLREEEQLLRREEQLLRQERDRKLRE 744

Query: 338 --QIMKETEASYKAQONLYIFLENIDRLQSLRLQAWTDKQKGLEEKHRECL 385
 Q+++E+E + E +L+ R + + +++ L+E+ E L

Sbjct: 745 EEQLLQSEERLRQ-----EREQQLRRERDRKFREEEQLLQREERL 789

Score = 103 (15.5 bits), Expect = 1.2e-01, P = 1.1e-01
 Identities = 42/152 (27%), Positives = 74/152 (48%)

Query: 36 ERPVAESLGHKDKDQEDYFQKGGQLKIFHCSKQLSLESSRQVTSSESQEEPWEEFG-REM 94
 ER + K +++E ++ +++ ++L E + + EQE E + RE

Sbjct: 835 ERLRRQERERKLREEEQLLRQEEQLRQERARKLR-EEEQLLRQEEQLRQERDRKLREE 893

Query: 95 RRQLWLEEEEMWQQRQKWA----LLEQEHQEKLRQWNLEDLAREQ---RRWVQ-LEKE 146
 + L EE+E+ Q+R +K LL++ +E+LR+ E RE++ RR Q L +E

Sbjct: 894 EQLLRQEEQLRQERDRKLREEEQLLQSEERLRQERERKLREEEQLLRREEQLRRE 953

Query: 147 QESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
 + RE EQL ++ E R R L + E

Sbjct: 954 RARKLREEEQLLQEREEERLRRQERARKLREEE 986

Score = 103 (15.5 bits), Expect = 7.8e-01, P = 5.4e-01
Identities = 31/91 (34%), Positives = 52/91 (57%)

Query: 67 KQLSLESSRQVTSESQEEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQ 126
++L E R++ E Q EE+ R+ R+ EEE++ Q+R+++ L QE KLR+
Sbjct: 642 QELRQERERKLREEEQLLRREEQELRQERERKLREEEQLLQEREE-RLRRQERARKLRE 700

Query: 127 WNLEDLAREQRRWVQLEKEQESPRREPEQL 157
E L R+++ +L +E+E RE EQL
Sbjct: 701 E--EQLLRQEEQ--ELRQERERKLREEEQ 726

Score = 101 (15.2 bits), Expect = 2.0e-01, P = 1.8e-01
Identities = 38/111 (34%), Positives = 57/111 (51%)

Query: 72 ESSRQVTSESQEEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 130
E R++ E Q EE E RE R+L EEE++ Q+R+++ L QE KLR+ +
Sbjct: 931 ERERKLREEEQLLRREEQELRRERARKL-REEEQLLQEREE-RLRRQERARKLREEE-Q 987

Query: 131 DLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSL 182
L RE+Q +L +E++ RE EQL ++ E R R + E L
Sbjct: 988 LLRREEQ---ELRQERDRKFREEEQLLQEREEERLRRQERDRKFREEERQL 1035

Score = 101 (15.2 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 33/108 (30%), Positives = 56/108 (51%)

Query: 72 ESSRQVTSESQEEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 131
E R++ E Q EE+ R+ R+ EEE++ +Q +++ L QE KLR+ E
Sbjct: 841 ERERKLREEEQLLRREEQELRQERARKLREEEQLLRQEEQ---LRQERDRKLREE--EQ 895

Query: 132 LAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
L R+++ +L +E++ RE EQL ++ E R R L + E
Sbjct: 896 LLRQEEQ---ELRQERDRKLREEEQLLQSEEEERLRRQERERKLREEE 940

Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
Identities = 32/97 (32%), Positives = 50/97 (51%)

Query: 72 ESSRQVTSESQEEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 131
E R+ E Q EE E R L EEE Q +++ L QE + KLR+ E
Sbjct: 578 EKRRRQERERQYRELELRQEEQLRDRKLREEEQLLQEREEERLRRQERERKLREE--EQ 635

Query: 132 LAREQ-----QRRWVQLEKEQESPRREPEQLGEDVERRI 165
L R++ Q R +L +E++ RRE ++L ++ ER++
Sbjct: 636 LLRQEEQELRQERERKLREEEQLLRREEQELRQERERKL 674

Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
Identities = 34/111 (30%), Positives = 58/111 (52%)

Query: 67 KQLSLESSRQVTSESQ--EEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKL 124
++L E R++ E Q +E EE R+ R+ EEE++ +Q +++ L QE + KL
Sbjct: 664 QELRQERERKLREEEQLLQEREEERLRRQERARKLREEEQLLRQEEQ---LRQERERKL 720

Query: 125 RQWNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEK 177
R+ +L RE+Q L +E++ RE EQL ++ E R + L +
Sbjct: 721 REEE-QLLRREEQL---LRQERDRKLREEEQLLQSEEEERLRRQEREQQLRR 768

Score = 98 (14.7 bits), Expect = 2.6e+00, P = 9.2e-01
Identities = 37/146 (25%), Positives = 77/146 (52%)

Query: 20 EPVLLPLVDRRFPPKWERPVAESLGHKDKQEDYFQKGLQIKFHCSKQLSLESSRQVTS 79
E LL ++ ++ ER + E + +E+ ++ K +QL + +++
Sbjct: 655 EEQLLRREEQELRQERERKLREEEQLLQEREEERLRRQERARKLREEEQLLRQEEQELRQ 714

Query: 80 ESQEEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE-LAREQQR 138
E + + EEE + +RR+ L +E ++ +++ LL++ +E+LR+ E L RE+ R
Sbjct: 715 ERERKLREEE--QLLRREEQLRQERDRKLREEEQLLQSEEEERLRRQEREQQLRRERDR 772

Query: 139 RWVQLEKEQESPRREPEQLG-EDVERRI 165
++ E+EQ RE E+L ++ ER++
Sbjct: 773 KF--REEEQLLQEREEERLRRQERERKL 798

Score = 97 (14.6 bits), Expect = 3.3e+00, P = 9.6e-01
Identities = 38/129 (29%), Positives = 63/129 (48%)

Query: 72 ESSRQVTSESQ--EEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 129
E R++ E Q +E EE R+ R+ EEE++ +Q +++ L QE KLR+
Sbjct: 817 ERERKLREEEQLLQEREEERLRRQERERKLREEEQLLRQEEQ---LRQERARKLREE-- 871

Query: 130 EDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSLVPAPSRT 189
E L R+++ +L +E++ RE EQL E+ + R R L + E L+

Sbjct: 872 EQLLRQEEQ---ELRQERDRKLRREEEQLLRQEEQEL--RQERDRKLRREEE-QLLQESEEE 925

Query: 190 QSAHQSRPHL 200
+ Q R L

Sbjct: 926 RLRRQERERKL 936

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
Identities = 41/132 (31%), Positives = 69/132 (52%)

Query: 46 KDKDQEDYFQKGGGLQI-KFHCSKQLSLESSRQVTSSESQEEPWEEFGRMRRQLWLEEEE 104
+++ QE F + Q + ++QL E S Q E + E+ G+ R QL +EE
Sbjct: 473 RERRQELQFLEEEELQRRERAAQQLQEEDSFQEDRERRRRQEQRPQTWRWQL---QEE 529

Query: 105 MWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRRWVQLEKEQESPRPEQLGEDVERR 164
++R +A Q QE+LR+ E+L RE++R+ E+E+E E Q ED +RR
Sbjct: 530 AQRARRHTLYAKPGQ--QEQLREE--EELQREKRRQ----EREREYREEEKLRQEEDEKRR 581

Query: 165 IFTPTSRWRDLEK 177
++R+LE+
Sbjct: 582 RQERERQYRELEE 594

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
Identities = 35/138 (25%), Positives = 76/138 (55%)

Query: 28 DRFFPKKWERPVAESL-GHKDKQEDYFQKGGGLQIKFHCSKQLSLESSRQVTSSESQEPW 86
+R++ + E E L K +++E Q+ + ++ L Q+ + ++E
Sbjct: 586 ERQRELEELRQEEQLRDRKLRREEEQLLQEREERLRQERERKLRREEEQLLRQEEQEL-L 644

Query: 87 EEEFGRMRRQLWL---EEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRRWVQL 143
+E R++R + L EE+E+ Q+R++K L +E Q L++ E L R+++ R +L
Sbjct: 645 RQERERKLRREEEQLLRREEQELRQERERK---LREEEQ-LLQEREERLRQERAR--KL 698

Query: 144 EKEQESPRPEPEQLGEDVERRI 165
+E++ R+E ++L ++ ER++
Sbjct: 699 REEEQLLRQEEQELRQERERKL 720

Score = 95 (14.3 bits), Expect = 5.2e+00, P = 9.9e-01
Identities = 59/282 (20%), Positives = 121/282 (42%)

Query: 20 EPVLLPLVDRFFPKKWERPVAESLGHKDKQEDYFQKGGGLQIKFHCSKQLSLESSRQVTS 79
E LL ++ ++ ER + E + +E+ ++ K +QL + +++
Sbjct: 655 EEQLLRREEQELRQERERKLRREEEQLLQEREERLRQERARKLRREEEQLLRQEEQELRQ 714

Query: 80 ESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLED-LAREQQR 138
E + + EEE + +RR+ L +E ++ +++ LL++ +E+LR+ E L RE+ R
Sbjct: 715 ERERKLRREE--QLLRREEQLLRQERDRKLRREEEQLLQESEERLRQEREQQLRRERDR 772

Query: 139 RWVQLEKEQESPRPEPEQLG-EDVERRIFTPTSRWRDLEKAELSLVPAPSTQSAHQ--S 195
++ E+EQ RE E+L ++ ER++ ++ E+ L + + Q
Sbjct: 773 KF--REEEQLLQEREERLRQERERKLRREEEQLLQEREERLRQERERKLRREEEQLLQ 830

Query: 196 RPHLPMPSTQFPALGKQRPMSVFTYRPTRRVPTKPKKSASFVPTGTSIRRLTWPS 255
R + ++ L ++ + E R R ++ +R+
Sbjct: 831 EREERLRQERERKLRREEEQLLRQE-EQLRQERARKLRREEEQLLRQEEQELRQERDRK 889

Query: 256 LQISPANIKKKVYHMDMEARQK---NLQLLSESELRLPHYLSKAL 299
L+ ++++ + E RK QLL E E RL R + L
Sbjct: 890 LREEEQLLRQEEQELRQERDRKLRREEEQLLQESEERLRQERERKL 936

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
Identities = 35/116 (30%), Positives = 59/116 (50%)

Query: 72 ESSRQVTSSESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK-----L 124
E +R++ E Q EE+ R+ R + + EEE++ Q+R+++ L QE K L
Sbjct: 977 ERARKLRREEEQLLRREEQELRQERDRKFRREEEQLLQEREEE-RLRRQERDRKFRREEERQL 1035

Query: 125 RQWNLEDLAREQQRRWVQLEKEQESPRPEPEQLGEDVERRIFTPTSRWRDLEKAELSL 182
R+ LE+ R+++ R +LE EQ +E +QL R F + R ++ E L
Sbjct: 1036 RQEELEEQFRQERDRKFRLE-EQIRQEEKQLRRQERDRKFRREEEQRRRQEREQQL 1092

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
Identities = 51/166 (30%), Positives = 76/166 (45%)

Query: 67 KQLSLESSRQVTSSESQ--EEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK 123
++L E R+ E Q +E EE R+ R R+L EEE++ + Q++ L QE+
Sbjct: 1250 QELRRERDRKFRREEEQLLQEREERLRQERARKLRREEEQLLFEEQEEQRL----RQER 1305

Query: 124 LRQWNLED-LAREQQRRWVQLEKEQESPRPEPEQLGEDVERRIFTPTSRWRDLEKAELSL 182
R++ E+ ARE++ R +LE+E R+E EQ R F R E+ E
Sbjct: 1306 DRRYRAEEQFAREKSR--RLEREL---RQEEEQRRRRERERKFRREEQRLRQEE-EQRR 1359

Query: 183 VPAPSRQSAHQSRPHLPMSPTQQPALGKQRPMSSEFTYRPRTRRPV 232
R QSRR L P T+Q A R E+ R++ P
Sbjct: 1360 RQLRRERQFREDQSRQVL--EPGTRQFARVPVRSSPLYEYIQEQRSQYRP 1407

Score = 93 (14.0 bits), Expect = 8.3e+00, P = 1.0e+00
Identities = 41/145 (28%), Positives = 72/145 (49%)

Query: 28 DRRFPKKWERPVAESLGHKDKDOEDYFQKGLQIKFHCSKQLSLESSRQVTSESQEEPW- 86
+RR ++ ER + E ++ Q + + Q + L R + QE+ +
Sbjct: 408 ERRQRQERERELEEQAARQQWQAEESERRRQ-RLSARPSLRERQLRAEERQEQRFR 466

Query: 87 -EEFGREMRRQL-WLEEEEMWQORQKKWALLEQE--HQEKLROWNLEDLAREQRRWVQ 142
EEE RE R++L +LEEEE Q+R++ L E++ +++ R+ ++ Q RW Q
Sbjct: 467 EEEQRRERRRQELQFLEEEELQRRERAQQLQEEDSFQEDRERRRRQEQRPQTWRW-Q 525

Query: 143 LEKEQESPRR----EP---EQLGEDVE 162
L++E + R +P EQL E+ E
Sbjct: 526 LQEEAQRRRHTLYAKPGQEQELREEEE 552

Score = 91 (13.7 bits), Expect = 2.4e+00, P = 9.1e-01
Identities = 38/110 (34%), Positives = 57/110 (51%)

Query: 72 ESSRQVTSESQEEPWEE-EFGREMRRQLWLEEEEMWQORQKKWALLEQEHOEKLROWN- 129
E R++ E Q EE E RE R+L EEE++ Q+R+++ L QE KLR+
Sbjct: 931 EREKRLREEEQLLRREEQELRRERARKL-REEEQLLQEREEE-RLRRQERARKLREEEQ 988

Query: 130 -----EDLAREQRRWVQLEKEQESPREPEQLGEDVERRIFTPTSRWRDLEKAE 180
++L +E+ R++ E+EQ RE E+L R F R L + EL
Sbjct: 989 LRREEQLRQERDRKF--REEEQLLQEREEERLRRQERDRKFREEER--QLRRQEL 1040

Score = 89 (13.4 bits), Expect = 2.2e+00, P = 8.9e-01
Identities = 35/138 (25%), Positives = 65/138 (47%)

Query: 82 QEEPWEEFGRMRRQLWLEEEEM--WQORQKKWALLEQEHOEKLROWNLEDLAREQRR 139
Q E++ E+R + + +E E WQ++++ L E+E Q K R+ + +R+ + +
Sbjct: 111 QNRRQEDQRRFELDRDQFEDEPERRRRQKQEERELAEERQKKRERFEQHYSRQYRDK 170

Query: 140 WVQLEKEQ-ESPREPEQL----GEDVERRIFTPTSRWRDLEKAELSLVPAPSRQSAHQ 194
+L+++ E R EQL G D E F + R E+ EL Q +
Sbjct: 171 EQRLRQLEERRAEELRRRGRDAEE--FIEEQLLRREQQELKR-ELREEEQRR 227

Query: 195 SRRPHLPMSPTQQPALGKQR 215
R H ++ L ++R
Sbjct: 228 RREQHERALQEEELRRQR 248

Score = 50 (7.5 bits), Expect = 2.2e+00, P = 8.9e-01
Identities = 34/160 (21%), Positives = 67/160 (41%)

Query: 325 RLQSLRQEAINHVOIMKETEASYKAQNLFIENIDRL-QSLRLQAWTDKQKGLEEKHRE 383
R + R+E Q+ +E E + + LE +R Q LR + +++ E++ R
Sbjct: 245 RQRRWREPREQQQLRRELEIREREQR---LEQERREQQLRREQRLEQERREQQLR 301

Query: 384 CLSSMVTMFPKLQLEWVHLNIP-EVTSPPKPKCKLPAASPRHIRPSGPTYKQPFSLRHR 442
L + +L+ E + E + K+L R R ++ L+
Sbjct: 302 ELEEREREQRLEQEERREQRLEQEERREQLKRELEIREREQRLEQEERREQLAEV 361

Query: 443 ACVPLQMARQOGKQMEAVWKTEVASSSYAIEKKTASLPRDQ 484
+ AR++G+ + W+ ++ S + A + K S PR Q
Sbjct: 362 R----EQARERGESLRRWQRQLESEAGARQSKV-YSRPRRQ 398

Score = 40 (6.0 bits), Expect = 1.9e-01, P = 1.7e-01
Identities = 32/115 (27%), Positives = 47/115 (40%)

Query: 276 RKNLQLLSESELRLPHYLRSKAL--ELTTTMMELGALRLQYLCHKYIFYRRL-QSLRQE 332
R+ QLL E E RL R++ L E E LR Q K+ +L Q +E
Sbjct: 959 REEEQLQEREEERLRRQERARKLREEEQLLRREEQELR-QERDRKFREEEQLLQEREEE 1017

Query: 333 AINHVOI---MKETEASYKAQNLFIENIDRLQSLRLQAWTDKQ-KGLEEKHRE 383
+ + +E E + Q L F + DR L Q +K+ K L + R+
Sbjct: 1018 RLRRQERDRKFREEERQLRRQLEEQFRQERDRKFLQEQIRQKEEKQLRRQERD 1073

Score = 37 (5.6 bits), Expect = 1.6e+00, P = 7.9e-01
Identities = 27/108 (25%), Positives = 43/108 (39%)

Query: 276 RKNLQLLSESELRLPHYLRSKAL--ELTTTMMELGALRLQYLCHKYIFYRRLQSLRQE 332
R+ QLL E E RL R+ L E E LR Q K R + L QE
Sbjct: 775 REEEQLQEREEERLRRQERARKLREEEQLLQEREEERLRRQERARKL---REEEQLLQE 831

Query: 333 AINHVOIMKETEASYKAQNLFIENIDRLQSLRLQAWTDKQKGLEEKHRE 383
+E E + + + E L+ R + +++ L ++ +E
Sbjct: 832 REEERLRRQERERKLREEEQLLRQEE-QELRQERARKLREEEQLLRQEEQE 881

Pedant information for DKFZphtes3_50n23, frame 1

Report for DKFZphtes3_50n23.1

[LENGTH] 499
[MW] 58885.69
[pI] 9.67
[KW] All_Alpha
[KW] LOW_COMPLEXITY 10.42 %

SEQ MTVRSRVADVFGSKDTESLEPVLLPLVDRRFPKKWERPVAESLGHKDKDQEDYFQKGGLO
SEG
PRD cccccceeeccccccccceeeccccccccccccchhhhhccccccccccccce

SEQ IKFHCSKQLSLESSRQVTSESQEEPWEEFEGREMRRQLWLEEEEMWQQRQKKWALLEQEH
SEG xxxxxxxxxxxx . xxxxxxxxxxxxxxxxxxxxxxxx
PRD eeeecchhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ QEKLRQWNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAEL
SEG
PRD hhh

SEQ SLVPAPSRTQSAHQSRRLPMPSPSTQQPALGKQRPMSVEFTYRPRTRRVPTKPKKSAS
SEG xxxxxxxxxxxxxxxxxxxxxxxx
PRD hccccccchhhhhccccccccccccccccccccccccceeeccccccccccccceee

SEQ FPVTGTSIRRLTWPSLQISPANIKKKVYHMDMEAQRKNLQLLSESELRLPHYLRSKALE
SEG xxxxxxxx
PRD eccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ LTTTMMELGALRLQYLCHKYIFYRRLQSLRQEAINHVVQIMKETEASYKAQNLYIFLENID
SEG
PRD hhh

SEQ RLQSLRLQAWTDKQKGLEEKHRECLSSMVTMFPKLQLEWNVHLNIPEVTSPPKPKCKLPA
SEG
PRD hhhhhhhhhhhhhcchhh

SEQ ASPRHIRPSGPTYKQPFSLRHRACVPLQMARQQGKQMEAVWKTEVASSSYAIEKKT PASL
SEG
PRD cccccccccccccchhhhhhhcchhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ PRDQLRGHPDIPRLTLTDV
SEG
PRD ccccccccccccccccccc

(No Prosite data available for DKFZphtes3_50n23.1)

(No Pfam data available for DKFZphtes3_50n23.1)

DKFZphtes3_6b21

group: testes derived

DKFZphtes3_6b21 encodes a novel 781 amino acid protein without similarity to human KIAA0256 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0256

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="356.3 cR from top of Chr9 linkage group"

Insert length: 3360 bp

Poly A stretch at pos. 3314, polyadenylation signal at pos. 3300

```

1  GGCAAGCCGA CGGCCCGCTG CTGGCCTCCG TGACGCGGCC TCCTCCGCGC
51  CTCGCGGCAT GCGCTCGGAG GGGCCGCGGG AGCCCGAAAG CGAGGGCATC
101 AAGTTATCAG CAGATGTCAA ACCATTGTGC CCCAGATTG CCGGGCTCAA
151 TGTGGCATGG TTAGAGTCCT CAGAAGCATG TGTCTTCCCC AGCTCTGCAG
201 CCACATACTA TCCGTTTGT TTTATAAACC AACCCAGTTG TTACCGAGGT
251 CAGTGCCTGG CTCCCAGTAT CTTTATAACC AACCCAGTTG TTACCGAGGT
301 TTTCAAACAG TGAAGCATCG AAATGAGAAC ACATGCCCTC TCCCACAAGA
351 AATGAAAGCT CTGTTTAAGA AGAAAACCTA TGATGAGAAA AAAACGTATG
401 ATCAGCAAAA GTTTGACAGT GAAAGGGCTG ATGGAATAT ATCATCTGAG
451 ATAAATATCAG CTAGAGGTTT ACATCATTTG TCCATTACG CTGAGAAATG
501 TTTGAAATCA GATGGTTACC ATAAGCGAAC AGACAGGAAA TCCAGAAATC
551 TTGCAAAAAA TGTATCTACC TCCAAACCTG AGTTTGAATT TACCACACTG
601 GACTTTCCTG AACTGCAAGG TGCAGAGAAC AATATGTGAG AGATACAGAA
651 GCAACCCCAAG TGGGGACCTG TCCACTCTGT CTCTACCGAC ATTCTCTTTC
701 TAAGAGAACT AGTAAACCCA GCTGCAGTGT TATCAAAGGG TGAATAGTGT
751 GTGAAAAATA ACCCAAATGA ATCTGTAATC GCTAATGCCG CTACCAATTC
801 TCCTTCATGT ACAAGAGAGT TATCTTGGAC ACCAATGGGT TATGTTGTTT
851 GACAGACATT ATCTACAGAA CTGTACAGCAG CCCCTAAAAA TGTACTTCTT
901 ATGATAAATC TAAAGACCAT TGCTTCATCA GCAGATCCTA AAAATGTTAG
951 TATACCATCT TCTGAAGCTT TATCTTCGGA TCCTTCTTAC AACAAAGAAA
1001 AACACATTAT TCATCTTACC CAAAAGTCTA AAGCATCACA AGGTAGTGAC
1051 CTTGAACAAA ATGAAGCCTC AAGAAAGAA AAGAAAAAGA AAGAAAAATC
1101 TACATCAAAA TATGAAGTCC TGACAGTTCA AGAGCCTCCA AGGATTGAAG
1151 ATGCCGAGGA ATTTCCCAAC CTGGCAGTTG CATCTGAAAG AAGAGACAGA
1201 ATAGAGACAC CGAAATTTC ATCTAAGCAG CAGCCACAGG ATAATTTTAA
1251 AAATAATGTA AAGAAGAGCC AGCTTCCAGT GCAGTTGGAC TTGGGGGGCA
1301 TGCTGACAGC CTGGGAGAAG AAGCAGCACT CTCAGCATGC AAAGCAGTCC
1351 TCCAAACCAAG TGGTAGTCTC AGTTGGAGCA GTGCCAGTGC TTTCCAAAGA
1401 ATGTGCATCA GGGGAGAGAG GCCGCCGCAT GAGTCAAATG AAGACCCCGC
1451 ACAATCCCTT GGACTCCAGC GCCCCACTGA TGAAGAAAGG GAAGCAGAGG
1501 GAGATCCCTA AGGCCAAGAA GCCAACCTCA CTGAAGAAGA TTATTTTGAA
1551 AGAACGGCAA GAGAGAAAGC AGCGTCTCCA AGAAAATGCT GTGAGTCCAG
1601 CTTTTACCAG TGATGACACA CAAGATGGAG AGAGTGGTGG TGATGACCAG
1651 TTTCCCGAGC AGGCAGAGCT GTCAGGGCCA GAGGGGATGG ACGAACTGAT
1701 CTCCACTCCT TCGGTTGAGG ACAAGTCTGA AGAGCCACCA GGCACAGAGC
1751 TCCAGAGGGA CACAGAGGCC TCCCACCTTG CTCCCAATCA CACCACCTTC
1801 CCTAAGATCC ACAGCCGCAG ATTCAGGGAT TACTGCAGCC AGATGCTTAG
1851 TAAAGAAGTG GATGCTTGTG TTACCGACCT ACTCAAAGAA CTGGTCCGTT
1901 TCCAAGACCG TATGTACCAG AAAGATCCAG TCAAGGCCAA GACTAAACGT
1951 CGACTTGTGT TGGGGTTGAG GGAGGTCTC AAACACCTGA AGCTCAAAAA
2001 ACTGAAATGT GTCAATTATT CTCCCAACTG TGAGAAGATA CAGTCAAAAG
2051 GTGGGCTGGA TGACACTTTG CACACAATTA TTGATTATGC CTGTGAGCAG
2101 AACATTCCCT TTGTGTTTGC TCTCAACCGC AAAGCTCTGG GGCAGTCTT
2151 GAATAAGGCA GTTCCTGTCA GTGTGGTGGG GATCTTCAGC TATGATGGGG
2201 CCCAGGATCA GTTCCACAAG ATGGTTGAGC TGACAGTGGC GGCCCGACAG
2251 GCGTACAGA CCATGCTGGA GAATGTGCAG CAGGAGCTGG TGGGAGAGCC
2301 CAGGCTCAG GCACCTCCCA GCCTACCCAC ACAGGGCCCC AGCTGCCCTG
2351 CAGAAGATGG CCCCCAGCC CTGAAAGAAA AAGAAGAGCC ACACTACATT
2401 GAAATCTGGA AAAACATCT GGAAGCATAC AGTGGATGTA CCCTGGAGCT
2451 AGAAGAATCC TTGGAGGCTT CAACCTCTCA AATGATGAAT TTGAATTTAT
2501 GAGAGTTCTT GCCTGTGTGT CTGTATTTT GGTAAAGAGG GGAGGTCTGA
2551 AAAAGACTTT GGGGCTTTTT CTCTGTTTT TCATGACAAT GTAATTTGTG
2601 TAACTGTTG ATCTGGAAAT TGATCAGCAT TAAAGGGCAC ATGAAGCAGT
2651 GTCTGCAGGC GTTCAGTGCT GCGGAGCCTG TTAAGGTCA CTCAGATGTG

```

```

2701 CAGGTGTTAA TCTTCTCTAA AAGCCTGGTT ATACAGCTCT GGCTTTCTGA
2751 GCACACTACG GATCTGGAAA ATACTGGAAA ATGTGATACT TAGAATACTT
2801 TGGCTGCTAA GGAACTTCC TCTCCATTGC AGAATAGCTG AGCCAAGTGA
2851 GTGAGTTTGC AGAAAGCAGG TGGTGAGCTC CTGCCTGCTG GAGGTTGCCA
2901 TGGAGGGCCA TTCCTGCCCG GCAACAGCAC CGTCCTGCAG GGAGCCACTT
2951 GGCAGAGGGG TGCAGGGCTG CTGGTGTCAG AGCAAGAGGG CTACAGGGAA
3001 AGGGCCCTTT CTCAGGGGAT GTAGCTTTTT TAAAAGATTT GGGAACACTT
3051 GGAGGATTG CTAAAATGAG CCTCAGAAGG AAAATTGGTT TTCTAACCTG
3101 TGACTTTTTG AAATGAATTA TTCCTTTCAG TCTTTATTTT TCAAAGAAAC
3151 AATGTGTATT GAAGTACCTA GATTTGTTTG ATAATCAACA AATCTTTCCT
3201 TTTTCAATGA ACATATCTG AATGTGGTTT CTGTCTTAGA CCAGGAGGAC
3251 AGAGTTTGCT TTCATATTTT CCCTGTAAGT AAGAGGGCTT ATTTATTTTA
3301 AATAAAGAGT AATTATTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAAAAAAAAA

```

BLAST Results

Entry HS773347 from database EMBL:
human STS WI-18160.
Score = 813, P = 2.9e-30, identities = 167/171

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 157 bp to 2499 bp; peptide length: 781
Category: similarity to known protein

```

1 MVRVLRSMCL PQLCSHILSV CSGTTSDRNV YSVPGSQYLY NQPSCYRGFQ
51 TVKHRNENTC PLPOEMKALF KKKTYDEKKT YDQKQFDSEK ADGTISSEIK
101 SARGSHHLSI YAENSLKSDG YHKRTDRKSR IIAKNVSTSK PEFETTLDF
151 PELQGAENNM SEIQKQPKWG PVHSVSTDIS LLREVVKPAA VLSKGEIVVK
201 NNPNESTAN AATNSPSCTR ELSWTPMGYV VRQTLSTELS AAPKNVTSMI
251 NLKTIASSAD PKNVSIIPSE ALSSDPSYNK EKHIHPTQK SKASQGSOLE
301 QNEASRKNKK KKEKSTSKYE VLTVQEPPI EDAAEFNLA VASERRDRIE
351 TPKFQSKQPP QDNFKNNVKK SQLPVQLDLG GMLTALEKKQ HSQHAQSSK
401 PVVSVGVAVP VLSKECASGE RGRMSQMKT PHNPLDSSAP LMKKGKQREI
451 PKAKKPTSLK KIILKERQER KQRLQENAVS PAFTSDDTQD GESGGDDQFP
501 EQAELSGPEG MDELSTPSV EDKSEEPFGT ELQRDTEASH LAPNHTTFFK
551 IHSRRFRDYC SQMLSKEVDA CVDLLKELV RFQDRMYQKD PVKAKTKRRL
601 VLGLREVLKH LKLKKLKCVI ISPNCIKQS KGGLDDTLHT IIDYACEQNI
651 PFVFALNRKA LGRSLNKAVP VSVVGIFSVD GAQDQFHKMV ELTVARQAY
701 KTMLENVQOE LVGEPRQPAP PSLPTQGPSC PAEDGPPALK EKEEPHYIEI
751 WKKHLEAYSG CTLELESLE ASTSQMMNLN L

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6b21, frame 1

SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256., N = 1, Score = 786, P = 3.6e-78

TREMBL:PFMAL3P3_15 gene: "MAL3P3.15"; Plasmodium falciparum MAL3P3, N = 2, Score = 161, P = 5.1e-10

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 150, P = 9.1e-07

>SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256.
Length = 635

HSPs:

Score = 786 (117.9 bits), Expect = 3.6e-78, P = 3.6e-78
Identities = 190/424 (44%), Positives = 263/424 (62%)

Query: 369 KKSQLPVQLDLGGMLTALEKKQHSQHAQ--SSKPVVSVGAVPVLSKECASGERGRMS 426
 KK++ PVQLDLG ML ALEK+Q + A+Q +++P+ +V + ++ + S
 Sbjct: 16 KKNKTPVQLDLGDLMAALEKQQAMKARQITNTRPLSYTVVTAASFHTKDSNTRKPLTKS 75

Query: 427 Q-MKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKERQERKQRLQENAVSPAFTS 485
 Q T N +D ++ KKGK++EI K K+PT+LKK+ILKER+E+K RL + S
 Sbjct: 76 QPCLTSFNSVDIASSKAKKGKEKEIAKLKRPTALKKVILKEREKKGRLTVD--HNLLGS 133

Query: 486 DDTQDGESGGDDQFPEQAELSGPEGMDLISTPSVEDKSEPPG--TELQDTEASHL-- 541
 ++ + D P++ G+ + S S+ S+ P T + + + AS
 Sbjct: 134 EEPTEMHLDIFIDDLPEIIVSQEDTGLS-MPSDTSLSPASQNSPYCMTFVSQSPASSGIG 192

Query: 542 APN-HTTFPKIHSRRFRDYCSQMLSKEVDACVTDLLKELVRFQDRMYQKDPVKAKTKRRL 600
 +P +T KIHS+RFR+YC+Q+L KE+D CVT LL+ELV FQ+R+YQKDPV+AK +RRL
 Sbjct: 193 SPMASSTITKIHSKRFRFCYCNQVLCKEIDECVTLLQLVLSFQERIYQKDPVRKARRRL 252

Query: 601 VLGLREV LKHLKLLKLCVVISPNCEKIQSKGGLDDTLHTIIDYACEQNIIPFVFALNRKA 660
 V+GLREV KH+KL K+KCVIISPNCCKIQSKGGLD+ L+ +I A EQ IPFVFAL RKA
 Sbjct: 253 VMGLREVTKMKLNKIKCVIISPNCCKIQSKGGLDEALYNVIAMAREQEIPFVFALGRKA 312

Query: 661 LGRSLNKAVPVSVVGIFSYDGAQDQFHKMVELTVAARQAYKTMLENVQOELVGEPRP--- 717
 LGR +NK VPVSUVGIF+Y GA+ F+K+VELT AR+AYK M+ ++QE E
 Sbjct: 313 LGRCVNKLVPVSUVGIFNYFGAESLFNKLVELTEEARAYKDMVAAMEQEQAEEALKNVK 372

Query: 718 QAPPSLP-TQGPS-----CPAEDGPPALKEKEEPHYIEIWKHLEAYSGCTL---ELE 766
 + P + ++ PS C P + E E Y W+ +E G E E
 Sbjct: 373 KVPHHMGRNPSAASAISSFCVISEP--ISEVNEKEYETNWRNMVETSDGLEASENEKE 430

Query: 767 ESLEASTSQ 775
 S + STS+
 Sbjct: 431 VSCKHSTSE 439

Pedant information for DKFZphtes3_6b21, frame 1

Report for DKFZphtes3_6b21.1

[LENGTH] 781
 [MW] 87393.44
 [pI] 8.94
 [HOMOL] SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256. 4e-75
 [PROSITE] MYRISTYL 4
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 16
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 16
 [PROSITE] ASN_GLYCOSYLATION 6
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 8.45 %

SEQ MVRVLRSMCLPQLCSHILSVCSGTTSDRNVYVPGSQYLYNQPSYRGFQTVKHRNENTC
 SEG
 PRD cccceeecc

SEQ PLPQEMKALFKKKTYDEKTYDQKFDSEADGTISSEIKSARGSHHLSIYAENSLKSDG
 SEGxxxxxxxxxxxxx.....
 PRD cccchhhhhhhhhccchhhhhhhhhccccchhhhhhhcccccccccccccccccccccccc

SEQ YHKRTDRKSRIIAKNVTSKPEFEFTTLDPELQGAENNMSEIQKQPKWGPVHSVSTDIS
 SEG
 PRD cccccchhhhhheccccccccccccccccccccccccccccchhhhhccccccccccccch

SEQ LLREVVKPAAVLSKGEIVVKNPNPESVTANAATNSPSCTRELSWTMGYVVRQTLSTELS
 SEG
 PRD hhhhhhheeecc

SEQ AAPKNVTSMINLKTIASSADPKNVSISSSEALSSDPSYNKEKHIHPTQKSKASQGSdle
 SEG
 PRD cccccccccchhhhhccch

SEQ QNEASRKNKKKKKSTSKYEVLTQVEPPRIEDAEFPNLAVASERRDIETPKFQSKQQP
 SEGxxxxxxxxxxxxx.....
 PRD hhhccccccccccccccccccccccccchhhhhccchhhhhhhhhhhcccccccccccc

SEQ QDNFKNNVKSQLPVQLDLGGMLTALEKKQHSQHAQSSSKPVVSVGAVPVLSKECASGE
 SEGxxxxxxxxxxxxxxxxxxxxx.....
 PRD cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccc

```

SEQ  RGRMSQMKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKERQERKQRLQENAVS
SEG  .....
PRD  chhhhhhhccccccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcc

SEQ  PAFTSDDTQDGESGGDDQFPEQAELSGPEGMDLISTPSVEDKSEEPGTELQRDTEASH
SEG  .....
PRD  cccccccccccccccccchhhhhhhccccceeecccccccccccccccccccccccccc

SEQ  LAPNHTTFPKIHSRRFRDYCSQMLSKEVDACVTDLLKELVRFQDRMYQKDPVKAKTKRRL
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhh

SEQ  VLGLREVLKHLKLLKLCVVISPNCEKIQSKGLDDTLHTIIDYACEQNIPFVFALNRKA
SEG  .....xxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhheeeccccccccccccchhhhhhhhhhhhhccccceeecccccc

SEQ  LGRSLNKAVPVSVVGIFSYDGAQDQFHKMVELTVAARQAYKTMLENVQQELVGEPRPOAP
SEG  .....
PRD  cccccceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

SEQ  PSLPTQGSPCAEDGFPALKEKEEPHYIEIWKHLEAYSGCTLEESLEASTSQMMNLN
SEG  .....xxxxxxxxxxxxx.....
PRD  cccccccccccccchhhhhhhccccceehhhhhhhhhchhhhhhhhhhhhhhhhhhhhhccccc

SEQ  L
SEG  .
PRD  C

```

Prosites for DKFZphtes3_6b21.1

PS00001	135->139	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	204->208	ASN_GLYCOSYLATION	PDOC00001
PS00001	245->249	ASN_GLYCOSYLATION	PDOC00001
PS00001	263->267	ASN_GLYCOSYLATION	PDOC00001
PS00001	544->548	ASN_GLYCOSYLATION	PDOC00001
PS00004	71->75	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	423->427	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	454->458	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	125->128	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	458->461	PKC_PHOSPHO_SITE	PDOC00005
PS00005	553->556	PKC_PHOSPHO_SITE	PDOC00005
PS00005	596->599	PKC_PHOSPHO_SITE	PDOC00005
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	384->388	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	640->644	CK2_PHOSPHO_SITE	PDOC00006
PS00006	702->706	CK2_PHOSPHO_SITE	PDOC00006
PS00007	581->588	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	73->82	TYR_PHOSPHO_SITE	PDOC00007
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	633->639	MYRISTYL	PDOC00008
PS00009	421->425	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_6b21.1)

DKFZphtes3_6c11

group: signal transduction

DKFZphtes3_6c11 encodes a novel 1025 amino acid protein with similarity to *A. ambisexualis* antheridio1 steroid receptor.

The novel protein is a putative steroid receptor. It shares similarity with yeast YNL132w and contains the ATP/GTP-binding site motif A (P-loop) and RGD site, similar to the *A. ambisexualis* antheridio1 steroid receptor.

The new protein can find application in modulating/blocking the expression of genes controlled by this receptor.

strong similarity to YNL132w

strong similarity to *S.pombe*/YDK9_SCHPO, *S.cerevisiae*/YNL132w, *C.elegans*/F55A12.8

Sequenced by BMFZ

Locus: unknown

Insert length: 3966 bp

Poly A stretch at pos. 3890, polyadenylation signal at pos. 3873

```

1 GCTGTGCCTT CTCTTTCGGA GTTGTTCCTG GCTCCACAGT GCTTCCCCTT
51 CTCCTACTGGC TGGGATCCCC CGGGCTCGGG GCGCAGTAAT AATTTTTCAC
101 CATGCATCGG AAAAAGGTGG ATAACCGAAT CCGGATTCTC ATTGAGAATG
151 GAGTAGCTGA GCGGCAAGA TCTCTCTTTG TTGTAGTTGG GGATCGAGGA
201 AAAGATCAGG TGGTAATACT TCATCACATG TTATCCAAAG CAACTGTGAA
251 GGCTCGGCCT TCAGTGCTGT GGTGTTATAA GAAAGAGCTG GGGTTTAGCA
301 GTCACCGGAA GAAAGAATG CGACAGCTGC AGAAGAAAAT AAAGAATGGA
351 ACACTGAACA TAAAGCAGGA CGACCCCTTT GAACTCTTCA TAGCAGCCAC
401 AAACATTCCG TACTGCTACT ACAACGAGAC CCACAAGATC CTGGGCAATA
451 CCTTCGGCAT GTGTGTGCTG CAGGATTTTG AAGCCTTAAC TCCAAACTTG
501 CTGGCCAGGA CTGTAGAAAC AGTGAAGGT GGTGGGCTAG TGGTCATCCT
551 CCTACGGACC ATGAACCTAC TCAAGCAATT GTACACAGTG ACTATGGATG
601 TGCATTCCAG GTACAGAACT GAGGCCCATC AGGATGTGGT GGAAGATT
651 AATGAAAGGT TTATTCTGTC TCTGGCTCTT TGTAAGAAAT GTCTCGTCAT
701 TGATGACCAG CTCAACATCC TGCCCATCTC CTCCCACGTT GCCACCATGG
751 AGGCCCTGCC TCCCAGACT CCGGATGAGA GTCTTGGTCC TTCTGATCTG
801 GAGCTGAGGG AGTTGAAGGA GAGCTTGCTG GACACCCAGC CTGTGGGTGT
851 GTTGGTGGAG TGCTGTAAGA CTCTAGACCA GGCCAAAGCT GTCTTGAAT
901 TTATCCAGGG CATCTCTGAA AAGACCCTGA GGAGTACTGT TGCACTCACA
951 GCTGCTCGAG GACGGGGAAT ATCTGCAGCC CTGGGATTGG CGATTGCTGG
1001 GGCGGTGGCA TTTGGGTACT CCAATATCTT TGTACCTCC CCAAGCCCTG
1051 ATAACCTCCA TACTCTGTTT GAATTTGTAT TTAAGGATT TGATGCTCTG
1101 CAATATCAGG AACATCTGGA TTATGAGATT ATCCAGTCTC TAAATCCTGA
1151 ATTTAAACAA GCAGTGATCA GAGTGAATGT ATTTCCAGAA CACAGGCGAA
1201 CTATTTCAGT TATACATCCT GCAGATGCTG TGAAGCTGGG CCAGGCTGAA
1251 CTAGTTGTGA TTGATGAAGC TGCCGCCATC CCCTCCCTCT TGGTGAAGAG
1301 CCTACTTGGC CCCTACCTTG TTTTCATGGC ATCCACCATC AATGGCTATG
1351 AGGGCACTGG CCGGTCACTG TCCCTCAAGC TAATTCAGCA GCTCCGTCAA
1401 CAGAGCGCCC AGAGCCAGGT CAGCACCACCT GCTGAGAATA AGACCACGAC
1451 GACAGCCAGA TTGGCATCAG CGCGGACACT GCATGAGGTT TCCCTCCAGG
1501 AGTCAATCCG ATACGCCCTT GGGGATGCAG TGGAGAAGTG GCTGAATGAC
1551 TTGCTGTGCC TGGATTGCCT CAACATCACT CGGATAGTCT CAGGCTGCCC
1601 CTTGCCGTGA GCTTGTGAAC TGTACTATGT TAATAGAGAT ACCCTCTTTT
1651 GCTACCACAA GGCCTCTGAA GTTTCTCTCC AACGGCTTAT GGCCCTCTAC
1701 GTGGCTTCTC ACTACAAGAA CTCTCCCAAT GATCTCCAGA TGCTCTCCGA
1751 TGCACCTGCT CACCATCTCT TCTGCCTTCT GCCTCCTGTG CCCCCACCCC
1801 AGAATGCCCT TCCAGAGTG CTGTCTGTTA TCCAGGTGTG CCTTGAAGGG
1851 GAGATTTCCT GCCAGTCCAT CTTGAACAGT CTGTCTCGAG GCAAGAAGGC
1901 TTGAGGGGAC CTGATTCCAT GGACAGTGTC AGAACAGTTC CAAGATCCAG
1951 ACTTTGGTGG TCTGTCTGGT GGAAGGGTCG TTCGCATTGC TGTTCAACCA
2001 GATTATCAAG GGATGGGCTA TGGCAGCCGT GCTCTGCAGC TGCTGCAGAT
2051 GTACTATGAA GGCAGGTTTC CTTGTCTGGA GGAAGAGGTC CTTGAGACAC
2101 CACAGGAAAT TCACACCGTA AGCAGCGAGG CTGTCAAGTT GTTGAAGAG
2151 GTCATCACTC CCCGGAAGGA CCTGCCTCCT TTAATCTTCA AATTGAATGA
2201 GACGCTGCTC GAACGCTGCT ATTACCTGGG TGTTCTCTAT GGCTTGACCC
2251 CCAGGCTCCT CAAGTTCTGG AACGAGCTG GATTGTGTCC TGTTTATCTG
2301 AGACAGACCC CGAATGACCT GACCGGAGAG CACTCGTGCA TCATGCTGAA
2351 GACGCTCACT GATGAGGATG AGGCTGACCA GGGAGGCTGG CTTGCAGCCT
2401 TCTGGAAGA TTTCCGACGG CGGTCTCTAG CCTTGCTCTC CTACCACTTC
2451 AGTACCTTCT CTCCTTCCCT GGCTCTGAAC ATCATTGAGA ACAGGAACAT
2501 GGGGAAGCCA CCCCAGCCTG CCCTGAGCCG GGAGGAGCTG GAAGCACTCT

```

```

2551 TCCTCCCTTA TGACCTGAAG CGGCTGGAGA TGTATTCACG GAATATGGTG
2601 GACTATCACC TCATCATGGA CATGATCCCG GCCATCTCTC GCATCTATTT
2651 CCTGAACCAG CTGGGGGACC TGGCCCTGTC TGGCGCTCAG TCGGCTCTTC
2701 TCTTGGGGAT TGGCCTGCAG CATAAGTCTG TGGACCAGCT GGAAAAGGAG
2751 ATTGAGCTGC CCTCGGGCCA GTTGATGGGA CTTTTCACCC GGATCATCCG
2801 CAAAGTTGTG AAGCTATTTA ATGAAGTTCA GGAAAAGGCC ATTGAGGAGC
2851 AGATGGTGGC AGCGAAGGAT GTGGTCATGG AGCCACGAT GAAGACCCCTC
2901 AGTGACGACC TAGATGAAGC AGCAAAGGAA TTTCAGGAGA AACACAAGAA
2951 GGAAGTAGGG AAGCTGAAGA GCATGGACCT CTCTGAATAC ATAATCCGTG
3001 GGGACGATGA AGAGTGGAAT GAAGTTTGA ACAAAGCTGG GCCGAACGCC
3051 TCGATCATCA GCCTGAAAAG TGACAAGAAA AGGAAGTTAG AGGCCAAACA
3101 AGAACCCAAA CAGAGCAAGA AGTTGAAGAA CAGAGAGACA AAGAACAAAA
3151 AAGATATGAA ACTGAAGCGG AAGAAATAGT GAAGAGAAAC TCGGGCATCT
3201 GTGTTTGATC ATGGGAAGAT ACTCTCACTA ACTGAACCCCT CTCTGGCTGG
3251 ACTGTTAAAA GCAACGAGAG GCCCCGGCAC ACCTGGAAGC TGGCCGCGAA
3301 TTCGGCTCTT GGGCCTGTGT GTCTGTGAGC TCAACCTGGC TAAAGGCAGA
3351 GTCACCTCCA AATGGGTCTC TTTAGAACTT GATGGCTGGG CACTGCCATC
3401 TCTAGAATTG CCACGAGTCT CTCTCTTCCT GCCCAGTCCA GGGCCCTCCT
3451 TTCCTATAAG TTCATATTTT GCTTTGAGCC AGCTTTTGTG TCTCATTCCT
3501 ACACATGTGG AAGCCACGTT GCCTCTCGAC CGCCTGAGGC CCTTAAGTAC
3551 ATCGCTTTCT GGTGGTGCCC AGGAGGCTGC TGCTGGGCGG CTGGGTCTCT
3601 CTTTGTGGAC TTGTACCTGG AGCAGGAGGA ACTCCAGTCC GTCCCGGCAT
3651 CCATGGCAGC CCGCGGTTAG GTGCGCCAGG GTTTGCTGAT GTTGTCTTGT
3701 GCTGTTCCAC TCTTGGCTCC AGCAGACCCA CTGTCCCAGA AAAGCCTGAT
3751 CCTGTAGTTT ATGTAGAATG CCACATCTGC GTCCTCAAGA CCTGTTTCAT
3801 CCATTGGGGA AAAGATGTTG GGAAAGGCCA CTTTGCTCGC AGGGGTGAGG
3851 GGAAGGATAG AGAATCTATT TTTAATAAAT AACATTCTAG AATGAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 102 bp to 3176 bp; peptide length: 1025
 Category: similarity to unknown protein
 Classification: unclassified
 Prosite motifs: RGD (966-969)
 ATP_GTP_A (284-292)

```

1 MHRKKVDNRI RILIENGVAE RQSLFVVVG DRGKDQVVIL HHMLSKATVK
51 ARPSVLWCYK KELGFSSHRK KRMRLQKKI KNGTLNIKQD DPFELFIAAT
101 NIRYCYNET HKILGNTFGM CVLQDFEALT PNLLARTVET VEGGGLVVIL
151 LRTMNSLKQL YVTMDVHSR YRTEAHQDVV GRFNERFILS LASCKKCLVI
201 DDQLNILPIS SHVATMEALP PQTPDESIGP SDLELRELKE SLQDTQPVGV
251 LVDCCKTLDO AKAVLKFIGE ISEKTLRSTV ALTAARGRGK SAALGLAIAG
301 AVAFGYSNIF VTSPSPDNLH TLFEFVFKGF DALQYQEHLD YEIIQSLNPE
351 FNKAVIRVNV FREHRQTIQY IHPADAVKLG QAEVLVVIDEA AAIPPLPVKS
401 LLGPYLVFMA STINGYEGTG RSLSLKLIQO LRQSAQSQV STTAENKTTT
451 TARLASARTL HEVSLQESIR YAPGDAVEKW LNDLLCLDCL NITRIVSGCP
501 LPEACELYYV NRDTLFCYHK ASEVFLQRLM ALYVASHYKN SPNDLQMLSD
551 APAHHLFCLL PPVPTQNAL PEVLAVIQVC LEGEISRQSI LNSLSRGKKA
601 SGDLPWTVS EQFQDPDFGG LSGGRVVRIA VHPDYQGMGY GSRALQLLQM
651 YYEGRFPCLE EKVLETPQEI HTVSSEAVSL LEEVITPRKO LPPLLLKLNE
701 RPAERLDYLG VSYGLTPRL LKFWKRAFVP VYLRQTPNDL TGEHSCIMLK
751 TLTDEDEADQ GGWLAAFWKD FRRRFLALLS YQSTFSFSL ALNIIQNRNM
801 GKPAQPALSR EELEALFLPY DLKRLEMYSR NMVDYHLIMD MIPAISRIYF
851 LNQLGDLALS AAQSALLGI GLQHKSDVQL EKEIELPSGO LMGLFNRIIR
901 KVVKLFNEVQ EKAIEEQMVA AKDVVMEPTM KTLSDDLDEA AKEFQEKHKK
951 EVGKLSMDL SEYIIRGDDE EWNEVLNKG PNASIISLKS DKRRKLEAKO
1001 EPKQSKKLKN RETKNKDKM LKRRK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_6c11, frame 3

TREMBL:CEAF3130_4 gene: "F55A12.8"; *Caenorhabditis elegans* cosmid F55A12., N = 1, Score = 2782, P = 1.1e-289

PIR:S55151 probable membrane protein YNL132w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 2549, P = 3.5e-273

SWISSPROT:YXX1_ACHAM HYPOTHETICAL PROTEIN (FRAGMENT)., N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I., N = 1, Score = 2843, P = 3.8e-296

>SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I.

Length = 1,033

HSPs:

Score = 2843 (426.6 bits), Expect = 3.8e-296, P = 3.8e-296
Identities = 576/1033 (55%), Positives = 750/1033 (72%)

```

Query:      1 MHRKKVDNRIRILIENGVAERQSRSLFVVVGDRGKQDVVLHHMLSKATVKARPSVLWCYK 60
             M +K +D+RI LI+NG +E+QRS FVVVGDR +DQVV LH +LS++ V ARP+VLW YK
Sbjct:      1 MPKALDSRIPTLIKNGCQEQRSFFVVVGDRARDQVNLHWLLSQSKVAARPVNLWMYK 60

Query:     61 KEL-GFSSHRKKMRQLQKKIKNGTLNLIKQDDPFELFIAATNIRYCYYNETHKILGNTFG 119
             K+L GF+SHRKKR +++K+IK G + +DPFELF + TNIRYCY Y E+ KILG T+G
Sbjct:     61 KDLGFTSHRKKRENKIKKEIKRGIRDPSNEDPFELFCSITNIRYCYKRESEKILGQTYG 120

Query:    120 MCVLQDFEALTPNLLARTVETVEGGGLVILLRTMNSLKQLYTVTMDVHSRYRTEAHQDV 179
             M VLQDFEALTPNLLART+ETVEGGG+VV+LL +NSLKQLYT++MD+HSRYRTEAH DV
Sbjct:    121 MLVLQDFEALTPNLLARTIETVEGGGIVVLLHLKLSLKQLYTMDSMDIHSRYRTEAHSDV 180

Query:    180 VGRFNERFILSLASCKKCLVIDDQNLNIPISSHVATMEALPPQTPDESIGPSDLELRELK 239
             RFNERFILSL +C+ CLVIDD+LN+LPIS ++ALPP +++ + ++EL+
Sbjct:    181 TARFNERFILSLGNCENCLVIDDELNVLPISGG-KNVKALPPTLEEDN--STQNSIKELQ 237

Query:    240 ESLQDTPQVGVLDCCKTLDQAKAVLKFIIEGISEKTLRSTVALTAARGRGKSAALGLAIA 299
             ESL + P G LV KTLDQA+AVL F+E I EK+L+ TV+LTA RGRGKSAALGLAIA
Sbjct:    238 ESLGEDHPAGALVGVTKTLDQARAVLTFVESIVEKSLKGTVSLTAGRGRGKSAALGLAIA 297

Query:    300 GAVAFGYSNIFVTSPSPDNLHTLFEEFVKGFDAEQEHLDYEIIQSLNPEFNKAVIRVN 359
             A+A GYSNIF+TSPSP+NL TLFEE+FKGFDAL Y+EH+DY+IIQS NP ++ A++RVN
Sbjct:    298 AATAHGYSNIFITSPSPENLKTLEFIFKGFDAALNVEEHVDYDIIQSTNPAYHNAIVRVN 357

Query:    360 VFREHRQTIQYIHPADAVKLGQAEVLVVIDEAAAIPPLVKSLLGPYLVFMASTINGYEGT 419
             +FR+HRQTIQYI P D+ LGQAEVLVVIDEAAAIPPLV+ L+GPYLVFMASTINGYEGT
Sbjct:    358 IFRDHRQTIQYISPEDSNVLGQAEVLVVIDEAAAIPPLVRLIGPYLVFMASTINGYEGT 417

Query:    420 GRSLSLKLIIQLRQQAQSQVSTTAENKTTTARLASARTLHEVSLQESIRYAPGDAVEK 479
             GRSLSLKL+QQLR+QS S + NK+ + + + S RTL E+SL E IRYA GD +E
Sbjct:    418 GRSLSLKLIIQLRQQAQSQVSTTAENKTTTARLASARTLHEVSLQESIRYAPGDAVEK 474

Query:    480 WLNDLLCLDCLN-ITRIVS-GCPLPEACELYYVNRDTLFCYHKASEVFLQRLMALYVASH 537
             WLN LLCLD + ++R+ + G P P C LY V+RDTLF YH SE FLQR+M+LYVASH
Sbjct:    475 WLNKLLCLDAASYVSRMATQGFPHPECSLYRVSRTDLFSYHPISEAFLQRMMSLYVASH 534

Query:    538 YKNSPNDLQMSDAPAHHLFCLLPVPPTQNALPEVLAVIQVCLEGEISRQSIILNSLSRG 597
             YKNSPNDLQ++SDAPAH LF LLPPV LP+ + VIQ+ LEG ISR+SI+NSLSRG
Sbjct:    535 YKNSPNDLQMSDAPAHQLFVLLPPVDLKNPKLPDPICVIQLALEGSIRESIMNSLSRG 594

Query:    598 KKASGDLIPWTVSEQFDQDFGGLSGGRVVRIAVHPDYQGMGYGSRALQLLQMYEGRFP 657
             ++A GDLPW +S+QFQD +F L G R+VRIAV P++ MGYG+RA+QLL Y+EG+F
Sbjct:    595 QRAGGDLIPWLISQQFQDENFAALGGARIVRIAVSPEHVKMGYGTRAMQLLHEYFEGKFI 654

Query:    658 CLEEKVLETPQEIHVSSEAV---SLLEEIVTPR--KDLPPLLLKLNERPAERLDYLGVS 712
             E+ + + + E + +L E I R K +PPLLLKL+E E L Y+GVS
Sbjct:    655 SASEEFKAVKHSLKRIGDEEIENTALQTEKIHVRDARTMPLLLKLSELQPEPLHYVGVS 714

Query:    713 YGLTPRLKFWKRAGFVFPVYLQRTPNDLTGEHSCIMLKTLDDEADQGGWLAFAFKDFR 772
             YGLTP L KFWK R G+ P+YLRQT NDLTGEH+C+ML+ L D WL AF ++F
Sbjct:    715 YGLTPSLQKFWKREGYCPYLRLQRTANDLTGEHTCVMLRVLEGRDSE----WLGAFQNFY 770

Query:    773 RRFLALLSYQFSTFSPSLALNIIQNRNMGKP----AQPALSREELEALFLPYDLKRLEMY 828
             RRFL+LL YQF F+ AL+++ N G + L+ EE+ +F YDLKRLE Y
Sbjct:    771 RRFLSLLGYQFREFAAITALSVDACNNGTKYVVNSTSKLTNEEINNVEFESYDLKRLESY 830

```

```

Query:      829 SRNMVDYHLIMDMIPAISRIYFLNQLGD-LALSAAQSALLLGLGQLHKSVDQLEKEIELP 887
             S N++DYH+I+D++P ++ +YF + D + LS Q ++LL +GLQ+K++D LEKE LP
Sbjct:      831 SNNLLDYHVIVDLLPKLAHLYFSGKFDPVKLSPVQQSVLLALGLQYKTIIDTLEKEFNLP 890

Query:      888 SGQLMGLFNRIIRKVVKLNFNEVQEAKIEEMQAADVME-----PTMKTLSDDLDE 939
             S QL+ + ++ +K+K+ +E++ K IEE++ + K P ++L ++L E
Sbjct:      891 SNQLLAMLVKLSKKIMKCIDEIETKDIEELGSNKKTESSNSKLPFTPLQQSLEELQE 950

Query:      940 AAKEFQ-EKHKKEVGKLSKMDLSEYIIRGDDEEWNEVLNKGAGPNASIIISLKSDDRKRKLEA 998
             A E +K+ + ++DL +Y IRG++E+W KA N I R +
Sbjct:      951 GADEAMLALREKQRELINADLEKYAIRGNEEDW-----KAAEN-QIQKTNGKGARVVSI 1004

Query:      999 KQEPKQSKKL--KNRETKNKKMDMLKRRK 1025
             K E +++ L +++TK K K K +K
Sbjct:      1005 KGEKRKNNSLDASDKKTKKEKPSKKFRK 1033

```

Pedant information for DKFZphtes3_6c11, frame 3

Report for DKFZphtes3_6c11.3

```
[LENGTH]      1025
[MW]           115704.57
[pI]           8.50
[HOMOL]        PIR:S55151 probable membrane protein YNL132w - yeast (Saccharomyces cerevisiae)
0.0
[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YNL132w] 0.0
[FUNCAT]       r general function prediction [H. influenzae, HI1254] 2e-05
[PROSITE]      ATP_GTP_A 1
[PROSITE]      RGD 1
[KW]           Alpha_Beta
[KW]           LOW COMPLEXITY 11.80 %
```

```

SEQ      MHRKKVDNRIRILILIENGVAERQSLFVVVVDGRGKDQVVLHHLMSKATVKARPSVLWCYK
SEG      .....
PRD      cccccccchhhhhccccccccceeeeeeecccccceeeehhhhhhhhhhhcccceehhhh

SEQ      KELGFSSHRKMRQLQKKIKNGTLNLIKQDDPFELFIAATNIRICYYNETHKILGNTFGM
SEG      .....
PRD      hhhccccchhhhhhhhhhhhhhhhhccccccccceeeeccccceeeccccceccccce

SEQ      CVLQDFEALTPNLLARTVETVEGGGLVILLRTMNSLKQLYTVTMDVHSRYRTEAHQDVV
SEG      .....XXXXXXXXXXXXXXXXX.....
PRD      eehhhhhccccchhhhhhhhhccccceeeecchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      GRFNERFILSLASCKKCLVIDDQLNILPISSHVATMEALPPQTPDESIGSPDLELRELKE
SEG      .....
PRD      hhhhhhhhhhhccccceeeeeccceeeccccccccccccccccccccccccccccchhhhhhhhh

SEQ      SLQDTQPVGVLVDCCKTLDOAKAVLKFIEGISEKTLRSTVALTAARGRGKSAALGLAIAG
SEG      .....XXXXXXXXXXXXX.....
PRD      hccccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhecccccccchhhhhhhhhh

SEQ      AVAFGYSNIFVTSPPDNLHTLFEFVKFGFDALQYQEHLDYEIQLSNPEFNKAVIRVNV
SEG      xxx.....
PRD      hhhhhccccceccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccccccccceeeeh

SEQ      FREHRQTIQYIHPADAVKLGQAEVVIDEAAAIPLPLVKSLLGPYLVEMASTINGYEGTG
SEG      .....
PRD      hhhhhhheeeccccccccccccceeeehhhhhccccchhhhhhhccccceeeeecccccccccc

SEQ      RSLSLKLIQQLRQOQSAQSQVSTTAENKTTTARLASARTLHEVSLQESIRYAPGDAVEKW
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      cchhhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhhhhhhhhhceeeccccchhhh

SEQ      LNDLCLDCLNITRIVSGCPLPEACELYVYNRDTLFCYHKASEVFLQRLMALYVASHYKN
SEG      xxxxxxxxxxxx.....
PRD      hhhhhhccccceeeccccccccccccceeeccccccccccchhhhhhhhhhhhhhhhhhhcccc

SEQ      SPNDLQMLSADAPAHHLFCLLPVPPTQNALPEVLAVIQVCLEGEISRQSIWLSLGRGKKA
SEG      .....
PRD      cccccccccccccceeeecccccccccccchhhhhhhhhccccchhhhhhhhhcccccc

SEQ      SGDLPWTVSEQFQDPDFGGLSGGRVVRIVHVPDYQMGYGSRALQLLQMYEGRFPCLE
SEG      .....
PRD      cccchhhhhhhhhhhccccccccceeeecccccccccccchhhhhhhhhhhhhhhccccchhh

SEQ      EKVLETQEIHTVSSAEVSLLEEIVTPRKDLPPLLLKLNERPAPERLDYLGVSYGLTPRL

```

```

SEG .....xxxxxxx.....
PRD hhhhhccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccchhh

SEQ KFWKRAGFVPVYLRQTPNDLTGEHSCIMLKTLTDEDEADQGGWLAAFWKDFRRRFLALLS
SEG .....
PRD hhhhhccccccceccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhh

SEQ YQFSTFSPSLALNIIQNRNMGKPAQPALSREELEALFLPYDLKRLEMYSRNMVDYHLIMD
SEG .....
PRD hhhhcchhhhhhhhhhhccccccccchhhhhhhhhhhccccchhhhhhhhhccccchhhhhhh

SEQ MIPAIISRIYFLNQLGDLALSAAQSALLLGIGLQHKSVDOLEKEIELPSGQLMGLFNRIIR
SEG .....xxxxxxx.....
PRD hhhhhhhhhhhccccchhhhhhhhhhhccccchhhhhhhhhhhccccchhhhhhhhh

SEQ KVVKLFNEVQEKAIEEQMVAADVVMEPTMKTLSDDLDEAAKEFQEKHKKEVGKLSMDL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ SEYIIRGDDEEWNEVLNKGAPNASIISLKSDDKRLKLEAKQEPKQSKKLKNRETKNKKDMK
SEG .....xxxxxxx.....
PRD cccccccccchhhhhhhhhccccccccccccchhhhhhhhhccccccccccccccccchhh

SEQ LKRKK
SEG xxxxx
PRD hhccc

```

Prosite for DKFZphtes3_6c11.3

PS00016	966->969	RGD	PDOC00016
PS00017	284->292	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphtes3_6c11.3)

DKFZphtes3_6d16

group: testes derived

DKFZphtes3_6d16 encodes a novel 695 amino acid protein nearly identical to a sequence from human PAC Clone WUGSC:H_DJ1185I07.2.

The cDNA is different to the proposed gene model: it contains additional exons.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H_DJ1185I07.2, differences to genmodel

differences to genmodel of WUGSC:H_DJ1185I07.2 two exons skippt,

Sequenced by BMFZ

Locus: /map="7q11.23-q21"

Insert length: 4572 bp

Poly A stretch at pos. 4540, polyadenylation signal at pos. 4520

```

1  GGC GGC GCTA GCTTCGGAGT CTCCGCGCGC CACCTCAGCC GCCTCCTAGC
51  GGC GGC GGC GCTCCTCTAC GCCTAAAATG ACCAATGTGT GATTTCAGTG
101 GAATAAATGG CGTCCAAAGT CACAGATGCT ATAGTCTGGT ATCAAAAGAA
151 GATTGGAGCA TATGATCAAC AAATATGGGA AAAATCTGTT GAACAGAGAG
201 AAATCAAGGG GCTAAGGAAT AAACCAAGA AAACAGCACA TGTGAAACCA
251 GACCTCATAG ATGTTGATCT TGTAAGAGGG TCTGCATTTG CAAAGGCAAA
301 GCCTGAAAGT CCTTGGACTT CTCTGACCAG AAAGGGAATT GTTCGAGTTG
351 TATTTTTCCC CTTTTCTTC CCGTGGTGGT TACAAGTAAC ATCAAAGGTC
401 ATCTTTTCTT GGCTTCTTGT CCTTTATCTT CTCAAGTTG CTGCAATAGT
451 ATTATTCTGC TCCACTTCTA GCCCACACAG CATACTCTGT ACAGAGGTGA
501 TTGGGCCGAT ATGGCTGATG CTGCTCCTGG GAACGTGTGA TTGCCAGATT
551 GTTTCACAA GAACACCCAA ACCTCCTCTA AGTACAGGGG GTAAAGAAG
601 AAGGAAATTA AGAAAAGCAG CCCATTGGA AGTACATAGG GAAGGAGATG
651 GTTCTAGTAC CACAGATAAC ACACAAGAGG GAGCAGTTCA GAACCCAGGT
701 ACAAGCACCT CTCACAGCGT TGGCACTGTC TTCAGAGATC TCTGGCATGC
751 TGCTTTCTTT TTATCAGGAT CAAAGAAAGC AAAGAATTCA ATTGATAAAT
801 CAACGTAAAC TGACAATGGC TATGTATCCC TTGATGGGAA GAAGACTGTT
851 GAGGAAGGTG CTGAAACAGG ATACTCATTG CGTCGTCATG TGGACAGGAC
901 TCGACCAGAA GAGACAGCCT GGAACACAGG AACACTGAGG AATGGTCCCTA
951 GCAAAGATAC CCAAAGGACA ATAACAAATG TCTCTGATGA AGTCTCCAGT
1001 GAGGAAGGTG CTGAAACAGG ATACTCATTG CGTCGTCATG TGGACAGGAC
1051 TTCTGAAGGT GTTCTTCGGA ATAGAAAGTC ACACCATAT AAGAAACATT
1101 ACCCTAATGA GGACGCCCTT AAATCGGGTA CTAGTTGCAG CTCTCGCTGT
1151 TCAAGTTCCA GACAGGATTC TGAGAGTGCA AGGCCAGAA CTGAAACAGA
1201 AGATGTGTAA TGGGAAGACT TGTTACATTG TGCAGAATGC CATTTCATCT
1251 GTACCAGTGA GACAGATGTG GAAAATCATC AGATTAAATC ATGTGTGAAA
1301 AAAGAATATA GAGATGACCC TTTTCATCAG AGTCATTTGC CCTGGCTCCA
1351 TAGTTCCAC CCAGGATTAG AAAAAATAAG TGCTATAGTA TGGGAAGGTA
1401 ATGATTGTAA GAAAGCAGAC ATGTCTGTAC TTGAAATCAG TGGAAATGATA
1451 ATGAACAGAG TGAACAGCCA TATACCAGGA ATAGGATACC AGATTTTGG
1501 AAATGCAGTC TCTCTCATAC TGGGTTTAA TCCATTGTGT TTCCGACTTT
1551 CTCAGCTTAC AGACTTGGAA CAACTCACAG CACATTCTGC TTCAGAACTT
1601 TATGTGATTG CATTTGGTTC TAATGAAGAT GTCATAGTTC TTCTATGGT
1651 TATAATAAGT TTTGTGGTTC GCGTGTCTCT TGTGTGGATT TTCTTTTTTT
1701 TGCTCTGTGT AGCAGAAAGA ACTTATAAAC AGCGATTACT TTTTGCAAAA
1751 CTCCTTGGAC ATTTAACATC TGCAAGGAGG GCTCGAAAAT CTGAGGTTCC
1801 TCATTTCCGG TTGAAGAAAG TACAGAATAT AAAAATGTGG CTATCTCTCC
1851 GTTCTATCTT TAAGCGTCGA GGTCCCTCAG GATCAGTTGA TGTAATAGTT
1901 TCATCTGCTT TCTTATTGAC TATCTCAGTT GTATTTATCT GTTGTGCCCA
1951 GATAAACCTC TACTTGAAGA TGGAGAAAAA ACCTAACAAA AAGGAGGAAC
2001 TGACACTAGT GAATAATGTT TTAATACTGG CTACTAAACT GCTAAAGGAG
2051 TTGACAGTCT CTTTATGATT ATATGGGCTT ACAATGAATC CGCTGCTTTA
2101 TAACATCACC CAGGTTGTTA TCCTGTCAGC TGTTTCTGGT GTTATCAGTG
2151 ACTTGCTTGG ATTTAATTTA AAGCTATGGA AGATTAAGTC ATGACAATTC
2201 AAAGAAAAGA AGATGTAGCC TCTTTTCCAG AATAAGAGTA CTGACTAAGC
2251 TGCCTGAAAG CTTGTCACTG ATTCTTTGCT TCAGGAGTCT CAGCTAGGGA
2301 GTTGAAAGTG TTACATCAGA CTGTCTTGTG CAATTCTTAT ATTTATTTTA
2351 CTGGTTCACT TTTTATTACA TTTATTTTAG TCTTTATATT TTTATTTTAA
2401 AGCATTGATG TACTTAGTTG TTGAAAGGGT GATGAAACTG ATATCCAGAT
2451 ACTTGAGATC CTGGTAATTG GTCATAAATA ATTGGCAAAA TAACAAATTG
2501 TGAAAATAGA AGCCATTGCT CAGCACCGTT TCTCCATCAA TGCCGTGAAC
2551 TTGCCTTACT TGAGGAAAAA TTCTTTAACT TTGGAATATT GCATTGAACT
2601 CAGCTATACA CATAAACAT TTTCTTTGGT AAATCAAGAT CCAGTCAGGG

```

```

2651 TTTCTCTTGA ATTATTTTGG AACAAATGCCA GGATCCAAAC TGATTAAGTT
2701 ACAGTTTAAAG CACCCTTCAG TATTAATATA TACGGTATTA TATAACAGGT
2751 CAACAAGTGC TCCTTGATGA TAAAACTTGT AATAGAGCAA TAATTGTAAA
2801 TGGTTACCAT ACTGTAAGAT ATTTTGATAA AAATTAACATA GTAATACTTG
2851 TATTATTTTG AAACACTGGG CTGTTTGACAC AGCTCCAACCT GTGCATGCTC
2901 AAAATGTGCA CTTTTTAAAA TTGTTACTTT TAATGCGTAT CTTTATATGG
2951 GATCTGTTAT AGTATACTAG GGCATGATAT GGTATCCTTT TGAGTGAGGT
3001 ATATACTCAT CTCACAAGTG AAGTGCCTAC TGATATTACT AAAGTACATT
3051 ATGTTTACTC AAGTAAATAA TTTTCTCCCC ATGGTACACT CTAGTGTAGG
3101 CTATTCATAC CACACTGAAA TGAACAACCTG AAGAATAAGG CTAAGAACCA
3151 ATAAATATT TCTCTAATTG CTAGTTGTAA AACTGTATCC AAATTTTCAG
3201 AAAAGACAGC TTCAGCTTGC AAATTCTATC CTCTAAACTT ATCTGGTGCA
3251 TTCTCCCCAC CCCACCCCCA TTATATAAGG GCTATTTTAG ATGCTTTTAA
3301 CCTCCCCAAC AAATAATTTG CCAAGTGTCC AATGAGAACT TATCATGTTG
3351 GTGTGTTAGG TAAATCGGGC AAATATGATA GTGTCTTACA TTGGGCCTTG
3401 ATTTTAAGTT GTTATATTG TACAATCGAG TATTTTAGAA ATTACATGAA
3451 ACATGAAACA GTTTTGTCAA TTTTTTTAA ACTGGGCATC TGGTTTCTAA
3501 AAATTTTATT GAAACAATCT AGAATTTTCT TGGTGCAAAG TGTATCATGT
3551 GGAATATCCT CATATTTTAA CCATATTTTA AGAATTTTAA GACGATTAAT
3601 TGTAATAAT TTATTTGATT GGTGCAAGTC TAATCCCTAA ATCATAATCT
3651 TAAATACAGG AATGTGTGGA GAACAGAGCC ATGTCATATC ACTTTGCTCT
3701 TACCATTCTT TTTGATCAGC CTCAATTCAG CCTCATTTG TAGTATGTTT
3751 TTTCTTTCTA TGAAAAACAA CAGAAAGCAT TTCATTTTAT TTGCCTATGT
3801 TCAAAATATG TTAATAATGA CCAAAGTGCA TTCTGAGTTT TTTCAGGAA
3851 TGTAATACTG GAGCTTTAAG AACATACTTA GTTCTCATG TGAAAACTTA
3901 GGCTTTGTCT GATGTTTTTC CTTCCTCTAT TGTCTAATGT TGAGGTTGTT
3951 TTTAGGAATT ATGTTTATA AACTTTTCA ATATAAGGTA CATGCCTATA
4001 CAGAACTTAA CATTTTGAC AGAATATATC AAATATATT TGAGAAAAAA
4051 AGTACGGCAT GAGTTCTGTT AGGAATAAAA GATGAACTA TTGTATCTCA
4101 CAAAAAATCT TATTTAGAA TGGAATATT TTTGAGAAAA GTAGCTGAGT
4151 ATACTGGTTT AAGAAAAATG TGTGTTTGA TTAGGTTTAA CTTAGAGTTG
4201 GGAGTTGATT TATTAAGTAC AGTATACCTC TCAACAGTTT ATAAATAATA
4251 TGTGGAATTA TGTCAGTGTG GGCAGCAGTA GAATACTAAA AGGAAAATGT
4301 CATGTTAAGC AATTTAGAA CATTAACCTG ACTATTTTCA AAGCAGAAAA
4351 ATTGACATTG CTGCCTTTAA GAATACCATG AATGTAAGAA ATTGAAAGAA
4401 ATTGTAATAT ATCACAATA ATAGAAATCG CAGTTCAAAG AGAATGTGG
4451 CAGATGTTGT GTGTGAACCTG TTGTTTCTTT GCCACATGTG TTGTATTTGA
4501 AAGTTTACA GTAAAGTTTAA AATAAACAT TCTGTGACTG AAAAAAATAA
4551 AAAAAAATAA AAAAAAATAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 2191 bp; peptide length: 695

Category: known protein

Classification: unclassified

Prosites motifs: CYTOCHROME_C (375-381)

```

1 MASKVTDIAV WYQKKIGAYD QQIWEKSVEQ REIKGLRNKP KKTAAHVKPD
51 IDVDLVRGSA FAKAKPESPW TSLTRKGIVR VVFFPFFFRW WLQVTSKVIF
101 FWLLVLYLLQ VAAIVLFCST SSPHSIPLTE VIGPIWMLL LGTVHCQIVF
151 TRTPKPLPLT GKKRRRLRK AAHLEVHREG DGSSTTDNTQ EGAVQNHGTS
201 TSHSVGTVFR DLWHAFFLS GSKKAKNSID KSTETDNGYV SLDGKKT VKS
251 GEDGIONHEP QCETIRPEET AWNTGTLRNG PSKOTQRTIT NVSDEVSSSE
301 GPETGYSLRR HVDRTSEGLV RNRKSHHYKK HYPNEDAPKS GTSCSSRCSS
351 SRQDSSEARP ESETEDVLWE DLLHCAECHS SCTSETDVEN HQINPCVKKE
401 YRDPFHQSH LPWLHSSHPC LEKISAIWE GNDCKKADMS VLEISGMIMN
451 RVNSHIPGLG YQIFGNAVSL ILGLTPFVFR LSQATDLEQL TAHSASELYV
501 IAFGSNEDVI VLSMVIISFV VRVSLVWIFF FLICVAERTY KQRLLFKLF
551 GHLTSARRR KSEVPFRLK KVQNIKMWLS LRSYLKRRGP QRSVDIVVSS
601 AFLLTISVVF ICCAQINLYL KMEKKPNKKE ELTLVNNVLK LATKLLKELD
651 SPFRLYGLTM NPLLYNITQV VILSAVSGVI SLLGFNLKL WKIKS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6d16, frame 2

PIR:S38170 SRP40 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 100, P = 0.08

TREMBL:AC004990.1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone
DJ1185I07 from 7q11.23-q21, complete sequence., N = 2, Score = 2693, P
= 0

>TREMBL:AC004990.1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone
DJ1185I07 from 7q11.23-q21, complete sequence.
Length = 588

HSPs:

Score = 2693 (404.1 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
Identities = 510/515 (99%), Positives = 512/515 (99%)

```

Query:   35 GLRNKPKKTAHVKPDLDVLDVRGSFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV  94
          GLRNKPKKTAHVKPDLDVLDVRGSFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV
Sbjct:   1  GLRNKPKKTAHVKPDLDVLDVRGSFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV  60

Query:   95 TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTP  154
          TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTP
Sbjct:   61 TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTP  120

Query:  155 KPPLSTGGKRRRKLKAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH  214
          KPPLSTGGKRRRKLKAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH
Sbjct:  121 KPPLSTGGKRRRKLKAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH  180

Query:  215 AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHQPCEIRPEETAWNT  274
          AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHQPCEIRPEETAWNT
Sbjct:  181 AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHQPCEIRPEETAWNT  240

Query:  275 GTLRNGPSKDTQRTITNVSDVSSEEGPETGYSLRHHVDRTSEGLRNKSHHYKKHYPN  334
          GTLRNGPSKDTQRTITNVSDVSSEEGPETGYSLRHHVDRTSEGLRNKSHHYKKHYPN
Sbjct:  241 GTLRNGPSKDTQRTITNVSDVSSEEGPETGYSLRHHVDRTSEGLRNKSHHYKKHYPN  300

Query:  335 EDAPKSGTSCSSRCSSSRQDSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN  394
          EDAPKSGTSCSSRCSSSRQDSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN
Sbjct:  301 EDAPKSGTSCSSRCSSSRQDSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN  360

Query:  395 PCVKKEYRDDPFHQSHLPWLHSSHGPLEKISAIWEGNDCKKADMSVLEISGMIMNRVNS  454
          PCVKKEYRDDPFHQSHLPWLHSSHGPLEKISAIWEGNDCKKADMSVLEISGMIMNRVNS
Sbjct:  361 PCVKKEYRDDPFHQSHLPWLHSSHGPLEKISAIWEGNDCKKADMSVLEISGMIMNRVNS  420

Query:  455 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM  514
          HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM
Sbjct:  421 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM  480

Query:  515 VIISFVVRVSLVWIFFFLLCVAERTYKQRLLFACL  549
          VIISFVVRVSLVWIFFFLLCVAERTYKQ L+ K+
Sbjct:  481 VIISFVVRVSLVWIFFFLLCVAERTYKQINLYLKM  515

Score = 409 (61.4 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
Identities = 92/115 (80%), Positives = 98/115 (85%)

Query:  595 DVIVSS----AFLLTISVVF-----CCA-----QINLYLKMEKKPNKKEELTLVNNVLK  640
          DVIV S   +F++ +S+V+I   C A   QINLYLKMEKKPNKKEELTLVNNVLK
Sbjct:  474 DVIVLSMVIISFVVRVSLVWIFFFLLCVAERTYKQINLYLKMEKKPNKKEELTLVNNVLK  533

Query:  641 LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKWLKIKS  695
          LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKWLKIKS
Sbjct:  534 LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKWLKIKS  588

```

Pedant information for DKFZphtes3_6d16, frame 2

Report for DKFZphtes3_6d16.2

[LENGTH] 695
[MW] 78466.68
[pI] 9.30
[HOMOL] TREMBL:AC004990.1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone DJ1185I07
from 7q11.23-q21, complete sequence. 0.0

{PROSITE} CYTOCHROME_C 1
 {KW} TRANSMEMBRANE 6
 {KW} LOW_COMPLEXITY 5.32 %

```

SEQ  MASKVTDIAIVWYQKKIGAYDQOIWEKSVEQREIKGLRNKPKKTAHVKPDLDIDVDLVRGSA
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
MEM  .....

SEQ  FAKAKPESPWTSLTRGIVRVVFFFFFRWWLQVTSKVIFFWLLVLVLLQVAAIVLFCST
SEG  .....xxxxxxxxxxxxx.....
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  SSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTPKPLSTGGKRRRKLKAAHLEVHREG
SEG  .....xxxxxxxxxxxxx.....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  DGSSTTDNTQEGAVQNHTSTSHSVGTVFRDLWHAFFLSGSKKAKNSIDKSTETDNGYV
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  SLDGKKT VKSGEDGIQNHQP CETIRPEETA WNTGTLRNGPSKDTQRTITNVSDVSSEE
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  GPETGYSLRRHVDRTSEGLRNKRSHYKHYPNEDAPKSGTSCSSSRCSSSRQOSESARP
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  ESETEDVLWEDLLHCAECHSSCTSETDVENHQINPCVKKEYRDDPFHQSHLPWLHSSHPG
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  LEKISAIWEGNDCKKADMSVLEISGMIMNRVNSHIPGIGYQIFGNAVSLILGLTPFVFR
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  LSQATDLEQLTAHSASELYVIAFGSNEDVIVLSMVIISFVVRVSLVWIFFFLLCVAERTY
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  KQRLLFALFGLTSARRARKSEVPHFRLKKVQNIKMWLSLSYLRKRGQRSDVIVSS
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  AFLLTISVVFICCAQINLYLKMEKKPNKKEELTVNNVLKATKLLKELDSPFRLYGLTM
SEG  .....
PRD  eeeeeeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  NPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

```

Prosites for DKFZphtes3_6d16.2

PS00190 375->381 CYTOCHROME_C PDOC00169

(No Pfam data available for DKFZphtes3_6d16.2)

DKFZphtes3_72k11

group: testes derived
DKFZphtes3_72k11 encodes a novel 233 amino acid protein with similarity to S.pombe
hypothetical repeat-containing protein.

The novel protein contains 5 leucine zippers and a microbodies C-terminal targeting signal (S-K-L) signature. This sequence is responsible for transport of proteins from free polysomes into the microbodies.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.pombe hypothetical repeat-containing protein

complete cDNA, complete cds, 6 EST hits (3 from testis derived
libraries)

Sequenced by DKFZ

Locus: unknown

Insert length: 1134 bp

Poly A stretch at pos. 1124, polyadenylation signal at pos. 1088

```

1 AACCTTTCAA GTGCCCTC CTTTCCTTAA AGTCTTTAT AGGGGTCCCC
51 TTCTTGCCCA TCTCCATCCT GTGAGTCAGG ACTGAAAGGG CACAGACAGG
101 TCACTGCCAG CATTGTTGGG GCAAGCCTGC AAGCACGCAT CACTGGGGAT
151 CTGACATGAC AATGGCCGCC TGCCCCCTCT GAGGGCTACA GGAATTACCC
201 CAGTGGGAAG CAGCTAAGCA GGTCTGACCA GCCGACCTGG ACCTGGCCAA
251 GGGTCCTGTC ATCCCTCATG GCCACCCCGC CATTCCGGCT GATAAGGAAG
301 ATGTTTTCCT TCAAGGTGAG CAGATGGATG GGGCTTGCTT GCTTCCGGTC
351 CCTGGCGGCC TCCTCTCCCA GTATTGCGCA GAAGAACTA ATGCACAAGC
401 TGCAGGAGGA AAAGGCTTTT CGCGAAGAGA TGAAAATTTT TCGTGAAAAA
451 ATAGAGGACT TCAGGGAAGA GATGTGGACT TTCCGAGGCA AGATCCATGC
501 TTTCCGGGGC CAGATCCTGG GTTTTGGGA AGAGGAGAGA CCTTCTGGG
551 AAGAGGAGAA AACCTTCTGG AAAGAGGAAA AATCCTTCTG GGAATGGAA
601 AAGTCTTTCA GGGAGGAAGA GAAAACCTTC TGGAAAAAGT ACCGCACTTT
651 CTGGAAGGAG GATAAGGCCT TCTGGAAGA GGACAATGCC TTATGGGAAA
701 GAGACCGGAA CCTTCTTCAG GAGGACAAGG CCCTGTGGGA GGAAGAAAAG
751 GCCCTGTGGG TAGAGGAAAG AGCCCTCCTT GAGGGGGAGA AAGCCCTGTG
801 GGAAGATAAA ACGTCCCTCT GGGAGGAAGA GAATGCCCTC TGGGAGGAAG
851 AGAGGGCCTT CTGGATGGAG AACAATGGCC ACGTGCCGG AGAGCAGATG
901 CTCGAAGATG GGCCCCACAA CGCCAACAGA GGGCAGCGCT TGCTGGCCTT
951 CTCCCGAGGC AGGCGGTAGC CAGCATGCAG GTGAGGGCC CTGTGGTCCA
1001 GACTCCCTG GGTGGGATT CAAGTCCAGG GTGAGCCCAT GTGCTGGAGA
1051 AAATACACAC TCATTGGTCT CTTGCTTTG AAAGATCCAA TAAAGTCTGT
1101 AGGCAAGGTT TGGAAAACCA ACTTAAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 268 bp to 966 bp; peptide length: 233
Category: similarity to known protein
Prosite motifs: MICROBODIES_CTER (231-234)
LEUCINE_ZIPPER (142-164)
LEUCINE_ZIPPER (149-171)
LEUCINE_ZIPPER (156-178)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (170-192)
LEUCINE_ZIPPER (170-192)

BLASTP hits

Entry A45973 from database PIR:
trichohyalin - human
Score = 147, p = 3.0e-07, identities = 57/194, positives = 94/194

No Alert BLASTP hits found

Pedant information for DKFZphtes3_72k11, frame 1

Report for DKF2phtes3_72k11.1

```
SEQ      MATPPFRLIRKMFSEFKVSRWMLGACFRSLAASSPSIRQKKLMMHKLQEEKAFFREEMKIFRE
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhh

SEQ      KIEDFREEMWTFRGKIHAFRGQILGFWEERPFWEEEKTFWKEEKSFWEMEKSFREEEKT
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      hhhhhhhhhhhhhhhhhhhcCCCCCCCCCCCCCCCCCchhhhhhhhhhhhhhhhhhhhhhhh

SEQ      FWKKYRTFWKEDKAFWKEDNALWERDRNLLQEDKALWEEEKALWVEERALLEGEKALWED
SEG      .....
PRD      hhhhcCCCCCCCCchhhhhhhhhhhhhcHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

SEQ      KTSLWEEENALWEEERAFAWMENNGHVAGEQMLEDGPHNANRGQRALLAFSGRA
SEG      ..... xxxxxxxxxxxxxx...
PRD      cccchhhhhhhhhhhhhhhhhhhcCCCCCHHHHHHCCECCCECCCHHHHHHHHCCC
```

Prosites for DKFZphtes3_72k11.1

(No Pfam data available for DKFZphtes3_72k11.1)

DKFZphtes3_72k15

group: cell structure and motility

DKFZphtes3_72k15 encodes a novel 188 amino acid protein with strong similarity to *Rattus norvegicus* actin-filament binding protein Frabin.

FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus *fgd1* seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an *esin* yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus.

The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

strong similarity to actin-filament binding protein Frabin

2 EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1845 bp

Poly A stretch at pos. 1835, polyadenylation signal at pos. 1816

```
1 GTGATGGAGA GTGCTGTTAT GATAGATGAA TCTAGGAAAG CCTCTTTGGA
51 GATGTGATAC CTGAACAGAA CCCCGAATGA TAAGAAGAAA TACCACTGTT
101 TTAGGAGAGA TTGTCCTAAG CAGAGAACAG CAGTGCAAA GACCCCAAGA
151 CACATACACT TGGTTATTAA GAATGGGAGC AGCAAGGAGT ATGGCAAGAA
201 CACAGTGAGT TTTCCCTTGA GTGTGTGAGG AAGCCCTCAG AGTTTGTGAC
251 TGACTTTGAG AGGTTCTAGT GGAGGGGATC AGAGTGGAAA CAAAGAGACC
301 AGTTAAAAAG GTATGGCAGC ATGAATAAAA AAGTTTGTAG AGTATTTCATT
351 ATGCCTTCCA AATAAAAAAC TCTTTGGTTC ATAATTGTGT CATAAATTAA
401 GGACTGGCTA CACTGTACTA TTTAAAAATG TTAAGAAACA TCAATAAGTA
451 AAAATGTTAG GAAGAGATGA TAAATACGTA AGTATTATAT CTAACCTAAGT
501 CTTTACTAAC TAGTCACATT ATTAACAGT GCAAGGATCA AGAAAAGTTA
551 AGCGTTGAAA AATAAATAAA TAAGTTATAA ATAAAAATAA CAGCCCAAGG
601 AAATGTTCCA GTCCCCATAG GTAGACTCGG GGTCATCTTC TTTATTTAAA
651 TCTTTATTTA AATGTGGATA GCATCCCAAG AGACTTGGGT CTACACTAAG
701 AATATTCAAA TCCATGTTTC TGAACCATC AGAGATAGAA AAAAAAAGTA
751 GCGAATATCC CTTTCAACT GGAATAAACT TGTCTTAATT CTAGAACTTT
801 TCCATACCAA TGTTTTCATG CTCCTTTTGT ATTTTATCTT TTAGCTCATT
851 ATCAAATTAT AGTGATTTGA AGAAAGAGTC TGCTGTGAAC CTAATGCTC
901 CTAGAACCCC AGGAAGGCAT GGATTGACAA CCACACCTCA ACAAAAACTC
951 CTCTCCCAGC ACTTGCCACA GAGGCAGGGA AATGATACAG ATAAGACTCA
1001 GGGTGCACAG ACTTGTGTGG CCAACGGTGT AATGGCAGCA CAAAACCAGA
1051 TGGAATGTGA GGAGGAGAAA GCTGCCACTC TTAGCTCAGA TACTTCTATT
1101 CAAGCTTCTG AACCTTGCT TGATACGCAC ATAGTGAATG GAGAAAGAGA
1151 TGAAACTGCC ACAGCTCCTG CATCACCCAC AACAGATAGC TGTGATGGAA
1201 ATGCTTCTGA CAGTAGCTAC AGGACTCCAG GCATAGGCCC AGTGCTCCCC
1251 CTAGAAGAAA GAGGGGCAGA AACAGAAACC AAGGTACAAG AGAGGGAAAA
1301 TGGGGAAAGC CCTCTGGAAC TGGAGCAGCT GGACCAGCAC CATGAGATGA
1351 AGGTAGAGCA TGAGACTAGC TCATGAGCAG GGAAGCCCT GCCTATTCTGA
1401 TTGTTGTCTT AAAACTCTTT ATTTATTGCA CCCCTGAAAT GTATGAATCA
1451 GATCACCCAC ACTGGCAGTT AAACGATTTT CAAGCTCTGG CTGCTGATTA
1501 GCATTTCCCC TATGCTCTAA GCAGATATTT CACTTTTCTT TTTCTATGAG
1551 TTTCTGTATA TATCTCTGTT GTAATTTTCA GAGTCAGAAC AGTGTGGAAA
1601 CTTTAAATATA GGAATCCAC AAATGTATTG TTTTACATA GAAAGAAAAT
1651 GTTCCTTGTG GCTCTAGATG TTGGTGCTGT ATCCCTAATA CTTACGGGCC
1701 AAGCAAGAAG AAATTGTATA ATCTTTGTG TTCAGAAAGT TCTAATAGAA
1751 TAAATAGGCC TGTAAGATGA ACTTGCCACT AGTAAATGTT ACTTTTAAGG
1801 ACATGAATAT GGAAGTATTA AATTATTCAG CAGATAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

98334590:

Frabin, a novel FGD1-related actin filament-binding protein capable of changing cell shape and activating c-Jun N-terminal kinase.

Peptide information for frame 3

ORF from 810 bp to 1373 bp; peptide length: 188
Category: similarity to known protein
Classification: Cell structure/motility

```

1 MFSCFLCILS FSSLSNYSYL KESAVNLNA PRTPGRHGLT TTPQKLLSQ
51 HLPQRQGNLT DKTQGAQTCV ANGVMQAQNG MECEEEKAAT LSSDTSIQAS
101 EPLLDTHIVN GERDETATAP ASPTTDCDNG NASDSSYRTP GIGPVLPLEE
151 RGAETETKVQ ERENGESPLE LEQLDQHHEM KVEHETSS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72k15, frame 3

TREMBL:AF038388_1 product: "actin-filament binding protein Frabin";
Rattus norvegicus actin-filament binding protein Frabin mRNA, complete cds., N = 1, Score = 428, P = 1.8e-39

>TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus norvegicus actin-filament binding protein Frabin mRNA, complete cds.
Length = 766

HSPs:

Score = 428 (64.2 bits), Expect = 1.8e-39, P = 1.8e-39
Identities = 90/174 (51%), Positives = 115/174 (66%)

```

Query: 12 SSSLSNYSYDLKESAVNLNAPRTPGRHGLTTTPQKLLSQHLPQRQGNLTKTQGAQTCVA 71
      S LS+Y+D++K+S +NLN P+TP +HGLT+T QKL S PQ+Q D+D+ QG C+A
Sbjct: 31 SVLSSYTDVQKDSMTNLNIPQTPRQHGLTSTTPQKLP SHKSPQKQEKDSDQNGQGHGCLA 90

Query: 72 NGVMAAQNMCECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAPASPTTDCDNG 131
      NGV AAQ+QMECE EK A LS +T Q + D H++NG R+ET T AS T+S D N
Sbjct: 91 NGVAAAQSQMECEETEKAALSPETDTQTAAASPDHVLNGVRNETTDSASSVTNSHDEN 150

Query: 132 ASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEMKVEHE 185
      A DSS RT G LP +E E ++QERENG S L LDQHHE+K +E
Sbjct: 151 ACDSSCRTQGTDLGLPSKEGEPVIEAELQERENGLSTEGLNPLDQHHEVKETNE 204

```

Pedant information for DKFZphtes3_72k15, frame 3

Report for DKFZphtes3_72k15.3

```

[LENGTH] 188
[MW] 20388.32
[pI] 4.62
[HOMOL] TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus
norvegicus actin-filament binding protein Frabin mRNA, complete cds. 2e-38
[KW] All_Alpha
[KW] SIGNAL PEPTIDE 16
[KW] LOW_COMPLEXITY 12.77 %

```

```

SEQ MFSCFLCILSFSSLSNYSYDLKESAVNLNAPRTPGRHGLTTTPQKLLSQHLPQRQGNLT
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccchhhhhccccccccccccccccccccccccccccccccccccccccchhhhhhhcccccccc
SEQ DKTQGAQTCVANGVMQAQNMCECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAP

```



```
SEG .....xxxxx
PRD cccccceecchhhhhhhhhhhhhhhhhhhhhccccceecccccceeccccccccccc

SEQ ASPTTDSCDGNASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEM
SEG xxxxx.....
PRD cccccccccccccccccccccccccccccccccchhhhhhhhhccccchhhhhhhhhhh

SEQ KVEHETSS
SEG .....
PRD hhhhcccc
```

(No Prosite data available for DKFZphtes3_72k15.3)

(No Pfam data available for DKFZphtes3_72k15.3)

DKFZphtes3_72p16

group: intracellular transport and trafficking

DKFZphtes3_72p16 encodes a novel 796 amino acid protein with very strong similarity to Mus musculus maternal-embryonic 3 (Mem3) gene.

Mem3 was isolated from a partial subtraction library of mouse unfertilized eggs and preimplantation embryos. Its transcript is abundant in the unfertilized egg and also actively transcribed from the newly formed zygotic genome. As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP).

The new protein can find application in modulation the sorting of proteins into different compartments.

strong similarity to mouse MEM3 and yeast VPS35

Sequenced by DKFZ

Locus: /map="16p13.3"

Insert length: 2707 bp

Poly A stretch at pos. 2697, no polyadenylation signal found

```

1 CTACGCGCGG GCGGGGTGCT GCTTGCTGCA GGCTCTGGGG AGTCGCCATG
51 CCTACAACAC AGCAGTCCCC TCAGGATGAG CAGGAAAAGC TCTTGGATGA
101 AGCCATACAG GCTGTGAAGG TCCAGTCATT CCAAATGAAG AGATGCCTGG
151 ACAAAAACAA GCTTATGGAT TCTCTAAAAC ATGCTTCTAA TATGCTTGTT
201 GAACTCCGGA CTTCATGTT ATCACCAGAG AGTTACTATG AACTTTATAT
251 GGCCATTCTT GATGAAGTGC ACTACTTGGA GGTCTACCTG ACAGATGAGT
301 TTGCTAAAGG AAGGAAAGTG GCAGATCTCT ACGAACTTGT ACAGTATGCT
351 GGAACATTA TCCCAAGGCT TTACCTTTTG ATCAGAGTTG GAGTTGTATA
401 TGTCAAGTCA TTTCTCAGT CCAGGAAGGA TATTTTGAAA GATTTGCTAG
451 AAATGTGCGG TGGTGTGCAA CATCCCTTGA GGGGTCTGTT TCTTCGAAAT
501 TACCTTCTTC AGTGTAACAG AAATATCTTA CCTGATGAAG GAGAGCCAAC
551 AGATGAAGAA ACAACTGGTG ACATCAGTGA TTCCATGGAT TTTGTACTGC
601 TCAACTTTGC AGAAATGAAC AAGCTCTGGG TGCGAATGCA GCATCAGGGA
651 CATAGCCGAG ATAGAGAAAA AAGAGAACGA GAAAGACAAG AACTGAGAAT
701 TTTAGTGGGA ACAAAATTGG TGCGCCTCAG TCAGTTGGAA GGTGTAATATG
751 TGGAACGTGA CAAACAGATT GTTTTGACTG GCATATTGGA GCAAGTTGTA
801 AACTGTAGGG ATGCTTTGGC TCAAGAAATAT CTCATGGAGT GTATTATTCA
851 GGTTTTCCTT GATGAATTTC ACCTCCAGAC TTTGAATCCT TTTCTTCGGG
901 CCTGTGCTGA GTTACACCAG AATGTAAATG TGAAGAACAT AATCATTGCT
951 TTAATTGATA GATTAGCTTT ATTTGCTCAC CGTGAAGATG GACCTGGAAT
1001 CCCAGCGGAT ATTTAACTTT TTGATATATT TTCACAGCAG GTGGCTACAG
1051 TGATACAGTC TAGACAAGAC ATGCCTTCAG AGGATGTTGT ATCTTTACAA
1101 GTCTCTCTGA TTAATCTTGC CATGAAATGT TACCCTGATC GTGTGGACTA
1151 TGTGTGATAA GTTCTAGAAA CAACAGTGGA GATATTCAAT AAGCTCAACC
1201 TTGAACATAT TGCTACCAAGT AGTGCAAGTT CAAAGGAACT CACCAGACTT
1251 TTGAAAATAC CAGTTGACAC TTACAACAAT ATTTTAACAG TCTTGAAATT
1301 AAAACATTTT CACCCACTCT TTGAGTACTT TGACTACGAG TCCAGAAAGA
1351 GCATGAGTTG TTATGTGCTT AGTAATGTTT TGGATTATAA CACAGAAATT
1401 GTCTCTCAAG ACCAGGTGGA TTCCATAATG AATTTGGTAT CCACGTTGAT
1451 TCAAGATCAG CCAGATCAAC CTGTAGAAGA CCCTGATCCA GAAGATTTTG
1501 CTGATGAGCA GAGCCTTGTG GGCCGCTTCA TTCATCTGCT GCGCTCTGAG
1551 GACCTTGACC AGCAGTACTT GATTTTGAAC ACAGCACGAA AACATTTTGG
1601 AGCTGGTGGG AATCAGCGGA TTCGCTTCAC ACTGCCACCT TTGGTATTTG
1651 CAGCTTACCA GCTGGCTTTT CGATATAAAG AGAATTCTAA AGTGGATGAC
1701 AAATGGGAAA AGAAATGCCA GAAGATTTTTC TCATTTGCCC ACCAGACTAT
1751 CAGTGCTTTG ATCAAAGCAG AGCTGGCAGA ATTGCCCTTA AGACTTTTTC
1801 TTCAAGGAGC ACTAGCTGCT GGGGAAATTG GTTTTGAAAA TCATGAGACA
1851 GTCGCATATG AATTCATGTC CCAGGCATT TCTCTGTATG AAGATGAAAT
1901 CAGCGATTCC AAAGCACAGC TAGCTGCCAT CACCTTGATC ATTGGCACTT
1951 TTGAAGGATG GAAGTGCTTC AGTGAAGAGA ATCATGAACC TCTGAGGACT
2001 CAGTGTGCCC TTGCTGCATC CAAACTTCTA AAGAAACCTG ATCAGGGCCG
2051 AGCTGTGAGC ACCTGTGCAC ATCTCTTCTG GTCTGGCAGA AACACGGACA
2101 AAAATGGGGA GGAGCTTCAC GGAGGCAAGA GGGTAATGGA GTGCCTAAAA
2151 AAAGCTCTAA AAATAGCAAA TCAGTGATG GACCCCTCTC TACAAGTGCA
2201 GCTTTTATTA GAAATCTGTA ACAGATATAT CTATTTTAT GAAAAAGGAA
2251 ATGATGCGGT AACAAATCAG GTTTTAAACC AGCTTATCCA AAAGATTCGA
2301 GAAGACCTCC CGAATCTTGA ATCCAGTGAA GAAACAGAGC AGATTAACAA
2351 ACATTTTCAT AACACACTGG AGCATTGCGG CTGCGGCGG GAATCACCAG
2401 AATCCGAGGG GCCAATTTAT GAAGGTCTCA TCCTTTAAAA AGGAAATAGC
2451 TCACCATACT CCTTTCCATG TACATCCAGT GAGGGTTTTA TTACGCTAGG
2501 TTTCCCTTCC ATAGATTGTG CCTTTCAGAA ATGCTGAGGT AGGTTTCCCA

```

2551 TTTCTTACCT GTGATGTGTT TTACCCAGCA CCTCCGGACA CTCACCTTCA
 2601 GGACCTTAAT AAAATTATTC ACTTGGTAAG TGTCAAGTC TTTCTGATCA
 2651 CCCCAAGTAG CATGACTGAT CTGCAATTTA AAATTCCTGT GATCTGTAAA
 2701 AAAAAA

BLAST Results

Entry AC007225 from database EMBLNEW:
 Homo sapiens chromosome 16 clone 480G7, WORKING DRAFT SEQUENCE, 38
 unordered pieces.
 Score = 1081, P = 2.8e-217, identities = 219/221
 13 exons

Entry HS015146 from database EMBL:
 human STS WI-8848.
 Score = 2033, P = 2.9e-87, identities = 425/436

Medline entries

96327632:
 Genetic mapping and embryonic expression of a novel, maternally
 transcribed gene Mem3.

97258867:
 Endosome to Golgi retrieval of the vacuolar protein sorting receptor,
 Vps10p, requires the function of the
 VPS29, VPS30, and VPS35 gene products.

92360909:
 Alternative pathways for the sorting of soluble vacuolar proteins in
 yeast: a vps35 null mutant missorts and
 secretes only a subset of vacuolar hydrolases.

10198044:
 Distinct Domains within Vps35p Mediate the Retrieval of Two Different
 Cargo Proteins from the Yeast
 Prevacuolar/Endosomal Compartment

Peptide information for frame 3

ORF from 48 bp to 2435 bp; peptide length: 796
 Category: strong similarity to known protein
 Classification: unset

1 MPTTQSPQD EQEKLLDEAI QAVKVQSFQM KRCLDKNKLML DSLKHASNML
 51 GELRTSMLSP KSYLYLYMAI SDELHYLEVY LTDEFAGGRK VADLYELVQY
 101 AGNIIPRLYL LITVGVVYVK SFPQSRKDIL KDLVEMCRGV QHPLRGLFLR
 151 NYLLQCTRNI LPDEGEPTDE ETTGDISDSM DFVLLNFAEM NKLWVRMQHQ
 201 GHSRDREKRE RERQELRILV GTNLVRLSQL EGVNVERYKQ IVLTGILEQV
 251 VNCRDALAQE YLMECIIQVF PDEFHLQTLN PFLRACAEHL QNVNVKNIII
 301 ALIDRLALFA HREDGPGIPA DIKLFDFISQ QVATVIQSRQ DMPSEDVVSL
 351 QVSLINLAMK CYPDRVDYVD KVLETTVEIF NKLNLHIAT SSAVSKELTR
 401 LLKIPVDYTN NILTVLKLKH FHPLFEYFDY ESRKSMSCYV LSNVLDYNTE
 451 IVSQDQVDSI MNLVSTLIQD QPDQPVDPED PEDFADEQSL VGRFIHLLRS
 501 EDPDQQYLIL NTARKHFGAG GNQRIRFTLP PLVFAAYQLA FRYKENSQVD
 551 DKWEKKCQKI FSFAHQTISA LIKAEALFEL LRLFLQALA AGEIGFENHE
 601 TVAYEFMSQA FSLYEDEISD SKAQLAAITL IIGTFERMKC FSEENHEPLR
 651 TQCALAASKL LKKPDQGRAV STCAHLEWSG RNTDKNGEEL HGGKRVMECL
 701 KKALKIANQC MDPSLQVQLF IEILNRYIYF YEKENDAVTI QVLNQLIQKI
 751 REDLPNLESS EETEQINKHF HNTLEHLRLR RESPESEGPI YEGLLI

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72p16, frame 3

TREMBL:AF024504_3 gene: "A TM017A05.7"; Arabidopsis thaliana BAC
 TM017A05., N = 2, Score = 927, P = 1.9e-162

PIR:S56936 vacuolar protein-sorting protein VPS35 - yeast
(Saccharomyces cerevisiae), N = 3, Score = 826, P = 1.5e-116

TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
maternal-embryonic 3 (Mem3) mRNA, complete cds., N = 1, Score = 3376, P
= 0

TREMBL:S42186_1 gene: "VPS35"; product: "Vps35p"; VPS35=vacuolar
protein sorting [Saccharomyces cerevisiae=yeast, Genomic, 3790 nt], N =
3, Score = 813, P = 4.4e-115

>TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
maternal-embryonic 3 (Mem3) mRNA, complete cds.
Length = 754

HSPs:

Score = 3376 (506.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 666/721 (92%), Positives = 682/721 (94%)

```
Query:   78 EVYLTDEFAGKRKVADLYELVQYAGNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC 137
          +VYLTDEFAGK ++ADLYELVQY+GNIIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC
Sbjct:   34 KYVLTDEFAGKERLADLYELVQYSGNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC 93

Query:  138 RGQVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSMDFVLLNFAEMNKLWVRM 197
          RGQVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSMDFVLLNFAEMNKLWVRM
Sbjct:   94 RGQVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSMDFVLLNFAEMNKLWVRM 153

Query:  198 QHGHGSRDREKRERERQELRILVGTNLVRLSQLEG-VNVERYKQIVLTGILEQVVNCRDA 256
          QHGHGSRDREKRERERQELRILVGTNLV L+ + +QIVLTGILEQVVNCRDA
Sbjct:  154 QHGHGSRDREKRERERQELRILVGTNLVLTLSWRCKCGTLQQIVLTGILEQVVNCRDA 213

Query:  257 LAQEYLMECIIQVFPDEFHLQTLNPFRLRACAEHQNVNKNIIIALIDRLALFAHREDGP 316
          LAQE MECIIQVFPDEFHLQTLNPFRLRACAEHQNVNKNIIIALIDRLALFAHRE P
Sbjct:  214 LAQEISMECIIQVFPDEFHLQTLNPFRLRACAEHQNVNKNIIIALIDRLALFAHREMEP 273

Query:  317 GIPADIKLFDIFSQQVATVIQSRQDMPSEDVVSLSQVSLINLAMKCYPDRVDYVDKVLETT 376
          GIPA++KLFDIFSQQVATVIQSR+DMPSEDVVSLSQVSLINLAMKCYPDRVDYVDKVLETT
Sbjct:  274 GIPAEKLKFDIFSQQVATVIQSRDMPSEDVVSLSQVSLINLAMKCYPDRVDYVDKVLETT 333

Query:  377 VEIFNKLNLLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHFHPLFEYFDYESR--K 434
          VEIFNKLNLLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHFHPLFEYFDYES K
Sbjct:  334 VEIFNKLNLLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHFHPLFEYFDYESSPGK 393

Query:  435 SMSICYVLSNVLDYNTIEIVSQDQVDSIMNLVSTLIQDQDPQVEDPDPEDFADEQSLVGRF 494
          SMSICYVLSNVLDYNTIEIVSQDQVDSIMNLVSTLIQDQDPQVEDPDPEDFADEQSLVGRF
Sbjct:  394 SMSICYVLSNVLDYNTIEIVSQDQVDSIMNLVSTLIQDQDPQVEDPDPEDFADEQSLVGRF 453

Query:  495 IHLLRSEDPDQYLIILNTARKHFGAGGNQIRIFTLPLPLVFAAYQLAFRYKENSKVDDKWE 554
          IHLLRS+DPDQYLIILNTARKHFGAGGNQIRIFTLPLPLVFAAYQLAFRYKENS K
Sbjct:  454 IHLLRSDPDQYLIILNTARKHFGAGGNQIRIFTLPLPLVFAAYQLAFRYKENS KWTSGK 513

Query:  555 KKCQKIFSAHQTSALIKAEALPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY 614
          + ++ F HQTISALIKAEALPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY
Sbjct:  514 RNARRYFHLPHQTISALIKAEALPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY 573

Query:  615 EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTQCALAASKLLKKPDQGRAVSTCA 674
          EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRT+CALAASKLLKKPDQ C
Sbjct:  574 EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTECALAASKLLKKPDQAEREHMTCT 633

Query:  675 HLFWSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLFIEILNRYIIFYEKE 734
          L WSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLFIEILNRYIIFYEKE
Sbjct:  634 SL-WSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLFIEILNRYIIFYEKE 692

Query:  735 NDAVTIQVLNQLIQKIREDLNLESSEETEQINKHFHNTLEHLRLRRESPESEGPIYEG 794
          NDAVTIQVLNQLIQKIREDLNLESSEETEQINKHFHNTLEHLR RRESPESEGPIYEG
Sbjct:  693 NDAVTIQVLNQLIQKIREDLNLESSEETEQINKHFHNTLEHLRTRRESPESEGPIYEG 752

Query:  795 IL 796
          IL
Sbjct:  753 IL 754
```

Pedant information for DKFZphtes3_72p16, frame 3

Report for DKFZphtes3_72p16.3

[LENGTH] 796

[illegible]

```

SEG .....
PRD hhhhhhhhhhhccccceeeehhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhh
MEM .....

SEQ RESPESEGPIYEGIL
SEG .....
PRD hhccccccccceeeccc
MEM .....

```

(No Prosite data available for DKFZphtes3_72p16.3)

(No Pfam data available for DKFZphtes3_72p16.3)

DKF2phtes3_7b22

group: cell structure and motility

DKF2phtes3_7b22 encodes a novel 443 amino acid protein with weak similarity to paramyosins.

The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

similarity to paramyosins

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: /map="3"

Insert length: 2291 bp

Poly A stretch at pos. 2241, polyadenylation signal at pos. 2213

```

1 GGAAGAAAGG CTAGCGGGCG TTGGCCGTAT GTGGGTGTCT TGAGGCAGTT
51 TTTCAGTTCT TTCATTACC AAAGTGACAT GCACCTACTA GGTGCCAGGT
101 GTTTAGACGT ACATACAACC CTCTGCAAAA TCTTTCAGTG TAGTCTCTCG
151 TATGAAAAGT TTCCAGCCAA GAATTGCCAC TGCACCTGAG ATAAGGGGGA
201 TCCTGGCCAT TAAGGAAACC TTGCCTTCGA AACTGAGCCG TGAGGAACTA
251 TACAAAATGG GAAATTGGGA CAAATCCCGG TGGCTCATGA CACTAAGAAG
301 TAAAATTACG AACTCACTGA GCTGGAAGTC ATTCAACGGG AATTGAATAG
351 GTAACGACAC TTTTGTGAGA TTATAAATAT ACCACGGAGG GTACGGAAGC
401 TACAGAAGAA TGGAAAGAAG CAGCCTGGAA GACTCAAACC TTCCTCCAAA
451 AGTTTGGCAT TCTGAGATGA CGGTGTCACT GACAGGCGAA CCACCTAGTA
501 CCGTAGAAGA AGAAGGAATA CCTAAAGAAA CAGACATAGA AATCATCCCA
551 GAAATCCCGG AAACCTCTAG GCCACTGTCC CTTCCAGATG TGCTGAGGAT
601 CTCGGCAGTT CTGGAGGACA CCACAGACCA GCTCTCTATT CTGAACTACA
651 TCATGCCCGT TCAGTACGAA GGGAGACAGA GCATCTCGCT GAAAAGCAGA
701 GAAATGAATC TAGAAGGAAC GAATCTAGAC AAACCTCCAA TGGCCTCAAC
751 AATCACAAAA ATACCCAGTC CGTTAATAAC TGAGGAAGGA CCCAACTTGC
801 CAGAAATCAG ACACAGAGGC CGGTTCGCTG TGGAGTTTAA CAAAATGCAG
851 GATCTTGTCT TCAAAAACC TACAAGGCAG ACCATCATGA CTACGGAGAC
901 ACTGAAGAAA ATTCAGATTG ATAGGCAGTT TTTACGCGAT GTGATTGCAG
951 ATACCATTA GAGTTGCAA GATTCGGCCA CTTACAACAG TCTCTGCAA
1001 GCTTTGAGCA AAGAGAGGGA AAACAAAATG CATTCTATG ACATCATTGC
1051 CAGGGAGGAA AAAGGAAGAA AACAGATAAT ATCACTTCAA AAACAGCTAA
1101 TTAATGTCAA AAAGGAATGG CAATTTGAAG TCCAGAGTCA GAATGAGTAT
1151 ATTGCTAACC TCAAGGACCA ACTGCAAGAG ATGAAGGCAA AATCCAACTT
1201 GGAGAATCGC TACATGAAAA CCAATACCGA GCTGCAGATT GCCCAGACCC
1251 AGAAAAAGTG TAACAGAACA GAGGAACCTT TGGTGAAGA GATTGAGAAA
1301 CTCAGGATGA AAACCGAAGA AGAGGCCCGG ACTCATACAG AGATTGAAAT
1351 GTTCCTTAGA AAGGAGCAGC AGAAACTTGA GGAGAGGCTG GAGTCTGGA
1401 TGGAGAAATA CGATAAGCAC ACAGAAATGA AACAGAATGA ACTAAATGCT
1451 CTCAAAGCCA CAAAGGCCAG TGACTTAGCA CACCTTCAAG ACCTGGCAAA
1501 GATGATAAGA GAGTATGAAC AGGTCATCAT TGAAGATCGT ATAGAAAAGG
1551 AGAGGAGCAA GAAGAAGGTA AAACAGGATC TCTTGGAAAT AAAGAGCGTT
1601 ATAAAGCTCC AGGCCTGGTG GCGAGGCACT ATGATACGGA GAGAAATTGG
1651 TGGTTTCAAG ATGCCTAAAG ACAAAGTTGA TAGCAAGGAT TCAAAAGGCA
1701 AAGGTAAAGG CAAGGATAAG AGGAGAGGCA AGAAGAAGTG ACCAAGTTCT
1751 CTTTTGTGTT TTCTGCTGGT ATTCTGGAGG TGGGAAGGAC TTGGAGAGTT
1801 AAGAAACACC TGGTACCTCA AAGATGACTC ATCTACAGGT TGTTTCCTAT
1851 TGAGACTTTC CCAGGGAAGC CTGATTTTAC TTTGCCTGTT AATTTCACTC
1901 TGCCTGTTAG GTGGGTTTTC AAACCCTGAT TTAGGATTAC ACCATTGACT
1951 TAGGGCTTCC TCATACCTTG CTGGGAAGAA GTTCTAGTA GTCCTGTGAA
2001 GATTCAATTCT TCTTGCTCTT TCTCAGCAGA ACAAAGGAGT TCACTGGCTT
2051 AGCTACAGTG ACGCATTGAA ACTTGAGTAA TTCTGTAAAT GTCAGATTTT
2101 GATTTTACCC AATTTGTCTG TAGTGAAAAA ACTCTTATGA GCAAAAAGTAT
2151 TCAGTAGGAA TTACAATATG ATGTTATTAG CTGTCCAGCA TAATATATAC
2201 ACAGCAAAGT TTTAATAAAT GTTGGTTTCT GCCTGCCTTT TAAAAAATAA
2251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A

```

BLAST Results

Entry G36731 from database EMBL:
SHGC-52923 Human Homo sapiens STS cDNA.

Score = 2262, P = 1.3e-97, identities = 462/468

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 410 bp to 1738 bp: peptide length: 443
Category: similarity to known protein

```

1 MEEDSLEDN LPPKVHSEM TVSVTGEPPS TVEEGIPKE TDIEIPEIP
51 ETLEPLSLPD VLRISAVLED TTDQLSILNY IMPVQYEGRQ SICVKSREM
101 LEGTNLDKLP MASTITKIPS PLITEEGPNL PEIRHRGRFA VEFNKMQLDV
151 FKKPTRQTIM TTETLKKIQI DRQFFSDVIA DTIKELQDSA TYNLLQALS
201 KERENKMHFY DIIAREEKGR KQIISLQKQL INVKKWQFE VQSNEYIAN
251 LKDQLQEMKA KSNLENRYMK TNTELQIAQT QKKCNRTTEL LVEEIEKLRM
301 KTEEEARTHT EIEMFLRKEQ QKLEERLEFW MEKYDKDTEM KQNELNALKA
351 TKASDLAHLQ DLAKMIREYE QVIEDRIEK ERSKKVKQD LLELKSVIKL
401 QAWWRGTMIR REIGGFKMPK DKVDSKDSKG KGKGKDKRRG KKK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_7b22, frame 2

SWISSPROT:MYSP_BRUMA PARAMYOSIN., N = 1, Score = 158, P = 5.8e-08

PIR:A44972 paramyosin - nematode (*Dirofilaria immitis*) (fragment), N = 1, Score = 157, P = 7.1e-08

SWISSPROT:MYSP_ONCVO PARAMYOSIN., N = 1, Score = 157, P = 7.4e-08

PIR:S52537 emm L 15 protein - *Streptococcus pyogenes*, N = 1, Score = 151, P = 8.6e-08

>SWISSPROT:MYSP_BRUMA PARAMYOSIN.
Length = 880

HSPs:

Score = 158 (23.7 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 66/259 (25%), Positives = 125/259 (48%)

```

Query: 142 EFNKMDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIADTIKELQDSATYNLLQALSK 201
      + K + L K R T E K++ + +D +A + LQ A N LL+ +
Sbjct: 169 QLKKDKHLAEKAAERFEAQTVELSNKVEDLNHRVND-LAQQRQLQ--AENNDLLKEIHD 225

Query: 202 ER---ENKMHF-YDIIAREEKGRKQIISLQKQLINVKKWQFEVQSNEYIANLKQQLQE 257
      ++ +N H Y + + E+ R+++ +++ ++ + +VQ + + + D+ E
Sbjct: 226 QKVQLDNLQHVQYLAQQLLEEARRRLEDAERERSQLQAQLH-QVQLELDSVRTALDE--E 282

Query: 258 MKAKSNLENRYMKTNTTELQIAQTQKKCNRTTELLVEEIEKLRMKT-EEEARTHTIEMFL 316
      A++ E++ NTE I Q + K + L EE+E LR K +++A +IE+ L
Sbjct: 283 SAARAEAEHKLALANTE--ITQWKSFDAAEVALHHEEVEDLRKKMLQKQAEYEEQIEIML 340

Query: 317 RKEQQ--KLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAHLQDLAKMIREYEQVII 374
      +K Q K + RL+ +E D E QN + L+ K + L K + E + I
Sbjct: 341 QKISQLEKAKSRLQSEVEVLIVDLEKAQNTIAILERAK-----EQLEKTVNELKVRID 393

Query: 375 EDRIEKERSKKVKQDLELKSVIKL 400
      E +E E ++++ + L EL+ + L
Sbjct: 394 ELTVELEAAQREARAALAEQLKLN 419

```

Score = 118 (17.7 bits), Expect = 1.3e-03, P = 1.3e-03
Identities = 54/231 (23%), Positives = 108/231 (46%)

```

Query: 181 DTIKELQDSATYNLLQ---ALSKERENKMHFYDIIAREEKGR-KQIISLQKQLINVKK 235
      D +KE+ D LQ L+++ E + RE + Q+ +Q +L +V+
Sbjct: 218 DLLKEIHDQKVQLDNLQHVQYLAQQLLEEARRRLEDAERERSQLQAQLH-QVQLELDSVRT 277

```


Query: 236 EWQFE--VQSQNEY-IANLKDQLQEMKAKSNLENRYMKTNTE-LQIAQTQKKCNRTTELL 291
 E +++ E+ +A ++ + K+K + E E L+ QK+ E++
 Sbjct: 278 ALDEESARAEEAEHKLALANTEITQWKSKFDAEVALHHEEVEDLRKKMLQKQAEYEEQIE 337

Query: 292 VEEIEKLRLMKTEEEARTHTEIEMF---LRKEQQKLE--ERLEFWMEKYDKDTEMKQNELN 346
 + ++K+ + ++R +E+E+ L K Q + ER + +EK + +++ +EL
 Sbjct: 338 IM-LQKISQLEKAKSRLQSEVEVLIVDLEKAQNTIAILERAKEQLEKTVNELKVRIDELT 396

Query: 347 A-LKATKASDLAHLQDLAKMIREYEQVIIEDRIEKERSKKKVKQDLELKSIVI 398
 L+A + A L +L K+ YE+ + E + R KK++ DL E K +
 Sbjct: 397 VELEAAQREARAALAEQLKLNLYEKAV-EQKEALARENKKLQDDLHEAKEAL 448

Score = 107 (16.1 bits), Expect = 2.1e-02, P = 2.1e-02
 Identities = 49/279 (17%), Positives = 124/279 (44%)

Query: 123 ITEEGPNLPEIRHRGRFAV-EFNKMDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIAD 181
 I E L + R A+ E K+++L K ++ + E KK+Q D + +AD
 Sbjct: 392 IDELTVELEAAQREARAALAEQLKLNLYEKAVEQKEALAREN-KKLQDDLHEAKEALAD 450

Query: 182 TIKELQDSATYNSLLQALSKERENKMHFYDIIAREEKGRKQ--IISLQKQLINVKKQWQF 239
 ++L + N+ L +E + + + R+ + R Q + LQ+ I +++ Q
 Sbjct: 451 ANRKIHLDLENARLAGEIRELQTLAKESEAARRDAENRAQALAEQLQRIEMERRLOE 510

Query: 240 EVQSQNEYIANLKDQLQEMKAKSNLENRYMKTNTE-LQIAQTQKKCNRT-ELLVEEIEKL 298
 + + N++ ++ + A L + + E+ + + + E E+ V+ + +
 Sbjct: 511 KEEEMEALRKNMQFEIDRLTAA--LADAEARMKAEISRLKKKYQAEIAELEMTVDNLNRA 568

Query: 299 RMKTEEEARTHTEIEMFLRKEQQKLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAH 358
 ++ ++ + +E L+ + + +L+ +++Y + Q +++AL A + +
 Sbjct: 569 NIEAQTIKKQSEQLKILQASLEDTQRQLQOTLDQY----ALAQKVSALSA-ELEECKV 623

Query: 359 LQDLAKMIREYEQVIIEDRIEKERSKKKVKQDLELKSIVIQLQ 401
 D A R+ ++ +E+ + V +L +K+ ++ +
 Sbjct: 624 ALDNAIRARKQAEIDLEEANGRITDLVSVNNNLTAIKNKLETE 666

Pedant information for DKFZphtes3_7b22, frame 2

Report for DKFZphtes3_7b22.2

[LENGTH] 443
 [MW] 51917.95
 [PI] 6.18
 [HOMOL] PIR:S28589 trichohyalin - rabbit 2e-08
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-07
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-07
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 5e-06
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 3e-05
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 6e-05
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 6e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 08.16 extracellular transport [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 2e-04
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDL207w] 4e-04
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YDL207w] 4e-04
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL201c] 5e-04
 [EC] 3.6.1.32 Myosin ATPase 3e-08
 [PIRKW] phosphotransferase 6e-06
 [PIRKW] citrulline 8e-06
 [PIRKW] tandem repeat 1e-07
 [PIRKW] heart 6e-06
 [PIRKW] polymorphism 4e-06
 [PIRKW] serine/threonine-specific protein kinase 6e-06
 [PIRKW] DNA binding 8e-08

PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	182->186	CK2_PHOSPHO_SITE	PDOC00006
PS00006	243->247	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	310->314	CK2_PHOSPHO_SITE	PDOC00006
PS00007	261->269	TYR_PHOSPHO_SITE	PDOC00007
PS00007	184->193	TYR_PHOSPHO_SITE	PDOC00007
PS00009	218->222	AMIDATION	PDOC00009
PS00009	439->443	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_7b22.2)

DKF2phtes3_7d17

group: testes derived

DKF2phtes3_7d17 encodes a novel 633 amino acid protein with weak similarity to human KIAA0454.

Pfam predicts a TNFR/NGFR cysteine-rich region.
No informative BLAST results; No predictive prosite or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0454

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3608 bp

Poly A stretch at pos. 3587, polyadenylation signal at pos. 3570

```
1 GGGAAAGTTAC GGGGAAGTCC ACCCAGCGTT TCTCAGGCAA TCTGAAGGCA
51 AATCCTGTTT AGACCCAGGC GAAGGTTTCT GGTGACCCAG GCTCTCACCA
101 GCCAATGTGC CCTTGCCGTC CTCCTGAGGG TATCTGGAGC TTCAGTGCTG
151 TGTGCTCTTG GCCTCCACAC TGGGGATGCC ACTGACTCCC ACTGTCCAGG
201 GCTTCCAGTG GACTCTCCGA GGCCCTGATG TAGAAACTTC CCCATTGCGT
251 GCACCAAGAG CAGCCTCACA TGGTGTGGGC CGACATCAAG AGCTGCGAGA
301 TCCAACAGTC CCTGGCCCA CTTCTTCTGC CACAAACGTC AGCATGGTGG
351 TATCTGCCGG CCTTGGTCC GGTGAGAAGG CAGAGATGAA CATTCTAGAA
401 ATCAACAAGA AATCGCGCCC CCAGCTGGCA GAGAACAAC AGCAGTTTCA
451 AAACCTCAAA CAGAAATGTC TTGTAATCA AGTGGCCTAC TTCCTGGCCA
501 ACCGGCAAAA TAATTACGAC TATGAAGACT GCAAAGACCT CATAAAATCT
551 ATGCTGAGGG ATGAGCGGCT GCTCACAGAA GAGAAGCTTG CAGAGGAGCT
601 CGGGCAAGCT GAGGAGCTCA GGCAATATAA AGTCCTGGTT CACTCTCAGG
651 AACGAGAGCT GACCCAGTTA AGGGAGAAGT TACAGGAAGG GAGAGATGCC
701 TCCCGCTCAT TGAATCAGCA TCTCCAGGCC CTCCTCACTC CGGATGAGCC
751 GGACAACTCC CAGGGACGGG ACCTCCGAGA ACAGCTGGCT GAGGGATGTA
801 GGCTGGCACA GCACCTCGTC CAAAAGCTCA GCCCAGAAAA TGATGACGAT
851 GAGGATGAAG ATGTTAAAGT TGAGGAGGCT GAGAAAGTAC AGGAATTATA
901 TGCCCCCAGG GAGGTGCAGA AGGCTGAAGA AAAGGAAGTC CCTGAGGACT
951 CACTGGAGGA GTGTGCCATC ACTTGTTCAA ATAGCCACCA CCCTTGTGAG
1001 TCCAACAGC CTTACGGGAA CACCAGAATC ACATTTGAGG AAGACCAAGT
1051 CGACTCAACT CTCATTGACT CATCCTCTCA TGATGAATGG TTGGATGCTG
1101 TATGCATTAT CCCAGAAAAT GAAAGTGATC ATGAGCAAGA GGAAGAAAAA
1151 GGGCCAGTGT CTCCAGGAA TCTGCAGGAG TCTGAAGAGG AGGAAGCCCC
1201 CCAGGAGTCC TGGGATGAAG GTGATTGGAC TCTCTCAATT CCTCTGCACA
1251 TGTCTGCCTC ATACAGTCT GACAGGAGCA CCTTTCACTC AGTAGAGGAA
1301 CAGCAAGTCG GCTTGGCTCT TGACATAGGC AGACATTGGT GTGATCAAGT
1351 GAAAAAGGAG GACCAAGAGG CCACAAGTCC CAGGCTCAGC AGGGAGCTGC
1401 TGGATGAGAA AGAGCCTGAA GTCTTGCAAG ACTCACTGGA TAGATTTTAT
1451 TCAACTCCTT TTGAGTACCT GGAAGTCCCT GACTTATGCC AGCCCTACAG
1501 AAGTGACTTT TACTCATTGC AGGAACAACA CCTTGGCTTG GCTCTTGACT
1551 TGGACAGAAAT GAAAAAGGAC CAAGAAGAGG AAGAAGACCA AGGCCACCA
1601 TGCCCCCAGG TCAGCAGAGA GCTGCCGGAG GTAGTAGAGC CTGAGGACTT
1651 GCAGGACTCA CTGGATAGAT GGTATTGAC TCCTTTTCACT TATCCAGAAC
1701 TGCCTGATTG ATGCCAGCCC TACGGAAGTT GCTTTTACTC ATTGGAGGAA
1751 GAACACGTTG GCTTTTCTCT TGACGTGGAT GAAATTGAAA AGTACCAAGA
1801 AGGGGAAGAA GATCAAAAGC CACCATGCCC CAGGCTCAAC GAGGTGCTGA
1851 TGGAAGCAGA AGAGCCTGAA GTCTTGCAAG ACTCACTGGA TAGATGTTAT
1901 TCGACTACTT CAACTTACTT TCAACTACAT GCCTCATTCG AGCAGTACAG
1951 AAGTGCCTTT TACTCATTG AGGAACAGGA CGTCAGCTTG GCCCTTGACG
2001 TGGACAATAG GTTTTACTT TTGACAGTGA TAAGGCACCA CTGGCCTTC
2051 CAGATGGGAG TCATATTCCC AACTTAAGCA GCCCTTACTA AGCTGAGAGA
2101 TGTCAATTGC GCAGGCAGGA CCTATAGGCA CATGTAGGTT TGAATGAAAC
2151 TGTAGTTCCC TTTGGAAGCC CAGTCATAGG ATGGGAAGT GGGCATGGCT
2201 CTATTCCCTAT TCTCAGACCA TGCCAGTGGC CACCTGTGCT CAGTCTGAAG
2251 ACGTTGGACC CAAGTTAGGT GTGACACGTT CACACGACTA TGTAGCACAT
2301 GCCGGGAGTG ATCTGCCAGA CATTCTAATT TGAACCAGAT ATCTCTGGGT
2351 AGCTACAAAG TTCCTCAGGG GTTTCATTTT GCAGGCATGT CTCTGAGCTT
2401 CTATACCTGC TCAAGGTCAG GTGCATCTTT GTGTTTAGCT CATCCAAAGG
2451 GTTTACCCTG GTTTCATTGA ACCTAACCCC ATTCTTTGTA TCTTCAGTGT
2501 TGGTTTGTGT TAGCTGATCC ATCTGTAACA CAGGAGGGAT CCTTGGCTGA
2551 GGATTGTATT TCAGAACAC TGACTGCTCT TGACAGTTGT TAACCCACTA
2601 GGCTCCTTTG AGTAGAGAAG CCATAGTCCT TCAGCCTCCA ATTGATATCA
2651 ATACTTAGGA AGACCACAGC TAGACGGACA AACAGCATTG GGAGGCCTTA
```

```

2701 GTCCTGCTCC TTTCAATTCC ATCCTGTAAA GAACAGGAGT CAGGAGCCGC
2751 TGGCAAGAGA CAGCATGTCA CCTGGGACTC TGCCAGTGCA GAATATGAAC
2801 AATGCCATGT TCTTGCAGAA AATGCTTAGC CTGAGTTTCA TAGGAGGTAA
2851 TCACCAGACA ACTGCAGAAT GTAGAACACT GAGCAGGACA ACTGACCTGT
2901 CTCCTTCACA CAGTCCACGT CACCACGAAT CACACAACAA AAAGGAGGAG
2951 AGATATTTTG GGTTCAGAAG AAGTAAATGA TAATGTAGCT ACATTTCCTT
3001 AGTTATTTTG AACCCCAAAT ATTTCCTCAT CTTTTTGTG TTGTCATTGA
3051 TTTTGGTGAC ATGGACTTGT TTGTAGAGGA CAGGTCAGCT GTCTGGCTCA
3101 ATGGTCTACA TTCTGAAGTT GTCTGAAAAT GTCTTCATGA TTAATTCAG
3151 CCTAAACGTT TCATCAAGAA CACTACAGAG TCGATACTGT GAGTTTCCAA
3201 CCTCAGCCCA TCTGTGGGCA GAGAAGGTCT AGTTTGTCCA TCAGCATTAT
3251 CATGATATCA GGA CTGGTTA CTTGGTTAAG GAGGGGTCTA GGAGATCTGT
3301 CCCTTTTAGA GACACCTTAC TTATGATGAA GTATTGGGGA GAGTGGTTTT
3351 TCAAAGTAGA AATGTCCTGT ATTCCAGTGA TCATCCTCTA AACGTTTTAT
3401 CATTATTATA TGCTGCCTCA ATGTTTACTG TGCCTTTGTT TTTGCTAGTG
3451 GCTGGAAATT TGCTGCCTCA ATGTTTACTG TGCCTTTGTT TTTGCTAGTG
3501 TGTGTTGTTG AAAAAAAAC ATTCTCTGCC TGAGTTTAA TTTTGTCCA
3551 AAGTTATTTT AATCTATACA ATTAAAACT TTTGCCTATC AAAAAAAAC
3601 AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 176 bp to 2074 bp; peptide length: 633
 Category: similarity to known protein

```

1  MPLTPTVQGF QWTLRGPDVE TSPFGAPRAA SHGVGRHQEL RDPTVPGPTS
51 SATNVSMVVS AGPWSGEKAE MNILEINKKS RPQLAENKQQ FRNLKQKCLV
101 TVQVAYFLANR QNNYDYEDCK DLIKSMRLDE RLLTEKLAE ELGQAEELRQ
151 YKVLVHSQER ELTQLREKLQ EGRDASRSLN QHLQALLTPD EPDNSQGRDL
201 REQLAEGCRL AQHLVQKLSF ENDDDEDEDV KVEEAQKVQE LYAPREVQKA
251 EEKEVPEDSL EECATCSNS HHPCESNQPY GNTRITFEED QVDSTLIDSS
301 SHDEWLDAVC IIPENESDHE QEEKEGVPSP RNLQSEEEEE APQESWDEGD
351 WTLISPPDMS ASYQSDRSTF HVSVEEQVGL ALDIGRHWCD QVKKEDQEAT
401 SPRLSRELLD EKEPEVLQDS LDRFYSTPFE YLELPDLCPQ YRSDFYSLQE
451 QHLGLALDLD RMKKDQEEEEE DQGPPCPRLS RELPEVVEPE DLQDSLDRWY
501 STPFSPYELP DSCQPYGSCF YSLEEHVGF SLDVDEIEKY QEGEEDQKPP
551 CPRLNEVLME AEEPEVLQDS LDRCYSTTST YFQLHASFQQ YRSFYSFEE
601 QDVSLALDVS NRFFTLTVIR HHLAFQMGVI FPH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7d17, frame 2

PIR:T00069 hypothetical protein KIAA0454 - human (fragment), N = 1,
 Score = 199, P = 1e-11

PIR:A45592 liver stage antigen LSA-1 - Plasmodium falciparum, N = 1,
 Score = 158, P = 2.7e-07

>PIR:T00069 hypothetical protein KIAA0454 - human (fragment)
 Length = 1,882

HSPs:

Score = 199 (29.9 bits), Expect = 1.0e-11, P = 1.0e-11
 Identities = 74/261 (28%), Positives = 122/261 (46%)

Query: 117 EDCKDLIKSMRLDERLLT----EELAEELGQAEELRQYKVLVHSQERELTQLREKLQEG 172
 +D + LI+ + + E L EEKLAEL A +Y L+ Q REL+ LR+K++EG
 Sbjct: 964 KDLESILQRVSQLAQLPKNGLEEKLAELRSASWPGKYDSLQDQARELSYLQKIREG 1023

Query: 173 RDAASRLNQH-----LQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDD 225
 R +H + LL ++ D G+ REQLA+G +L + L KLS ++
 Sbjct: 1024 RGICYLITRHAKTVKSFEDLLRSNDIDYYLGQSFREQLAQGSQQLTERLTSKLSTKDHKS 1083

Query: 226 EDEDVKVEEAQVQELYAPREVQKAEK-EVPEDSLEECITCSNSHHPCESNQPYGNTR 284
 E + +E L RE+Q+ E+ EV + L+ ++T S+SH +S++ +T
 Sbjct: 1084 EKDAQGLEPLA----LRLSRELQEKEKQVIEVLQAKLDARSLTPSSSHALSDSHRSPSSTS 1139

Query: 285 ITFEEDQV--DSTLIDSSSHDEWLDVAVCIIPENESDHEQEEKGPVSPRNQSEEEEEAP 342
 +E + D ++ +H E A P + +S + S + A
 Sbjct: 1140 FLSDELEACSDMDIVSEYTHYEKKAS---PSHSDSIHSSSHSAVLSSKPSSTSASQGA 1196

Query: 343 QESWDEGDWTLISIPDMSASYQSDRSTFH 371
 ES + +L P + S FH
 Sbjct: 1197 AES-NSNPISLPTPQNTPKANQAHSGFH 1224

Score = 89 (13.4 bits), Expect = 1.1e-01, P = 1.0e-01
 Identities = 35/89 (39%), Positives = 44/89 (49%)

Query: 464 KDQEEEEEDQG---PPCPRLSRELPEVVEP-EDLQDSLDRWYSTPFSYPELPDSCQ-PYGS 518
 KD + E+DQ P RLSREL E + E LQ LD TP S L DS + P +
 Sbjct: 1079 KDHKSEKDQAGLEPLALRLSRELQEKEKQVIEVLQAKLDARSLTPSSSHALSDSHRSPSSTS 1138

Query: 519 CFYSLEEEHVGFSLDVDEIEKYQEGEEDQKPP 550
 F S E E D+D + +Y EE + P
 Sbjct: 1139 SFLSDELEACS---DMDIVSEYTHYEKKASP 1167

Score = 73 (11.0 bits), Expect = 4.8e+00, P = 9.9e-01
 Identities = 31/88 (35%), Positives = 40/88 (45%)

Query: 390 DQVKKEDQEATSP---RLSRELLD-EKEPEVLQDSLDRFYSTPFEYLELPDLCQ-PYRSD 444
 D ++DQ P RLSREL + EK EVLQ LD TP S L D + P +
 Sbjct: 1080 DHKSEKDQAGLEPLALRLSRELQEKEKQVIEVLQAKLDARSLTPSSSHALSDSHRSPSSTS 1139

Query: 445 FYSLQEQHLGLALDLDRMKKQEEEEEDQGGP 475
 F S L D+D + + EE + P
 Sbjct: 1140 FLS---DELEACSDMDIVSEYTHYEKKASP 1167

Score = 68 (10.2 bits), Expect = 1.1e-01, P = 1.0e-01
 Identities = 36/156 (23%), Positives = 68/156 (43%)

Query: 31 SHGVGRHQELRDPTV---PGPTSSATNVSMVVSAGPWS-----GEKAEMNILEINKK 79
 S G +HQE + TV P P S + V A G ++ ++ +
 Sbjct: 684 SPGKHQHQEGEENVTVPFRPQSLDLGATFTVDAHQLDNQSQPRDPGPGSAFSLPGSTQH 743

Query: 80 SRPQLAENKQFRNLKQKCLVTQVAYFL-ANRQNNYDYE-DCKDLIKSMLRDERLLTEEK 137
 R QL++ KQ++++L++K L+++ F AN Y + L+K + ++ ++
 Sbjct: 744 LRSQLSQCKQRYQDLQEKLLSEATVFAQANELEKYRVMLTGESLVKQDSKQIQVDLQDL 803

Query: 138 LAEELGQAEELRQYKVLVHSQERELTQREK-LQEG 172
 E G++E + + + E L+E L EG
 Sbjct: 804 GYETCGRSENEAEREETTSPECEEHNSLKEMVLMG 839

Score = 65 (9.8 bits), Expect = 2.2e-01, P = 2.0e-01
 Identities = 23/96 (23%), Positives = 52/96 (54%)

Query: 123 IKSMRLDERLLTEEKLAELGQAE-----LRQYKVLVHSQERELTQREKLQEGRDASRS 178
 ++ + D+ + E + E+ EE LRQ ++ V ++ +L +LR+ L ++ +
 Sbjct: 5 LRQRIHDKAVALERAIDEKFSALEEKEKELRQLRLAVRERDHLERLRDVL-----SNEA 60

Query: 179 LNQHLLQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKL 218
 Q +++LL ++G ++ EQL+ C+ Q L +++
 Sbjct: 61 TMQSMESLL-----RAKGLEV-EQLSTTCQNQLWLKEEM 93

Score = 61 (9.2 bits), Expect = 5.5e-01, P = 4.2e-01
 Identities = 27/95 (28%), Positives = 47/95 (49%)

Query: 134 TEEK-LAEELGQAEELRQY---KVLVHSQERELTQREKLQEGRDASRSLNQHLQALLT 188
 +E K L +LG+ EE R Y +LV +++ L+ +LQ ++L +++L
 Sbjct: 855 SERKPLENLQKGQEEFRVYKSENILV--LRKDIKDLKAQLQANANKVIQNLKSRVRSLSV 912

Query: 189 PDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDDEDE 228
 + +S R R+ A G ++ SP + DEDE
 Sbjct: 913 TSDYSSSLERP-RKLRAVGT-----LEGSSPHSVPEDE 945

Score = 57 (8.6 bits), Expect = 1.4e+00, P = 7.5e-01
 Identities = 26/92 (28%), Positives = 47/92 (51%)

Query: 127 LRDERLLTEEKLAELGQAEEL---RQYKVLVHSQERELTQREKLQEGRDASRSLNQHL 183
 L E LL EK+A Q+E+ R+ ++L+ + L R +L E A R L L
 Sbjct: 358 LTQEVLLLRKVASVESQGEISGNRRQQLLLMLEG--LVDESRNLNEALQAERQLYSSL 415

Query: 184 QALLTPDEPDNSQ-GRDLREQLAEGCRLAQHLVQKL 218
 P++S+ R L+ +L EG ++ + +++
 Sbjct: 416 VKFHA--HPESSERDRTLQVEL-EGAQVLRSLREEV 448

Score = 54 (8.1 bits), Expect = 2.7e+00, P = 9.3e-01
 Identities = 61/264 (23%), Positives = 121/264 (45%)

Query: 3 LPTPTVQGFQWTLRGPVETSPFGAPRAASHGVGRHQE--LRDPTVPGPTSSATNVSMVVS 60
 L+ T Q QW L+ ++ET F + + + L D SAT ++
 Sbjct: 79 LSTTCQNQLW-LK-EEMETK-FSRWQKEQESIIQQLQTSLHNRNKEVEDLSAT---LLCK 132

Query: 61 AGPWSGEKAEMNILEINKKSR---PQLAENKQFRNLKQKCLVTQVAYFLANRQNNYDYE 117
 GP E AE + +K R L++ +Q L+ + + + ++ R+
 Sbjct: 133 LGPGQSEIAEELCQRLQKRMQLDQLLSDRNKQV--LEHEMEIQGLLSVSTREQE-SQA 189

Query: 118 DCKDLIKSMLRDERLLTEEKLAELGQAEELRQYKVLVHSQERELT---QLREKLQEG-- 172
 + L++++ ER + L + LG + L + + +Q+ E+T +L ++ +G
 Sbjct: 190 AAEKLVQALM--ERNSELQALRQYLGGDSLMS-QAPISNQAEVTPTRGLGKTQDQSM 246

Query: 173 RDASRSLNQHLLQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDDEDVVKV 232
 + SR + L A P ++ G DL + +A G L ++LS N +E E +
 Sbjct: 247 QIPSRDDSTSLTAKEDVSIPTSLG-DL-DTVA-G-----LEKELS--NAKEEELMAK 295

Query: 233 EEAKEVQELYAPREVQKAEKEVPEDSLEECAT 266
 +E E EL A + + +E+E+ + + ++T
 Sbjct: 296 KERESQMELSALQSMMAVQEEELQVQADMESLT 329

Score = 49 (7.4 bits), Expect = 6.3e+00, P = 1.0e+00
 Identities = 21/87 (24%), Positives = 39/87 (44%)

Query: 192 PONSQGRDLREQLAEGCRLAQHLVQKLSPENDDDDEDVKEEAEKVQELYAPREVQKAE 251
 P ++Q LR QL++ + Q L +KL + + E EK + + + K +
 Sbjct: 738 PGSTQ--HLRSQSLQCKQRYQDLQEKLLS---EATVFAQANELEKYRVMLTGESLVKQD 792

Query: 252 EKEVPEDSLEECAT-TCSNSHHPCESNQ 278
 K++ D L++ TC S + E +
 Sbjct: 793 SKQIQVD-LQDLGYETCGRSENEAEREE 819

Score = 46 (6.9 bits), Expect = 6.3e+00, P = 1.0e+00
 Identities = 19/77 (24%), Positives = 39/77 (50%)

Query: 112 NNYDYEDCKDLIKSMLRDERLLTEEKLAELGQAEELRQYKVLVHSQERELTQLREKLQ- 170
 + ++ E+ K + E ++T+E L+E QAE R+ + + + L+E+L
 Sbjct: 597 DGWEIEEDKE--KGEVMVETVVTKEGLSESSLQAE-FRKLQGLKNAHNIINLLKEQLVL 653

Query: 171 EGRDASRSLNQHLLQALLT 188
 ++ + L L LT
 Sbjct: 654 SSKEGNSKLTPELLVHLT 671

Pedant information for DKFZphtes3_7d17, frame 2

Report for DKFZphtes3_7d17.2

[LENGTH] 633
 [MW] 72951.15
 [pI] 4.40
 [HOMOL] PIR:T00069 hypothetical protein KIAA0454 - human (fragment) 2e-11
 [BLOCKS] BL00201E
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 14
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 2
 [PFAM] TNFR/NGFR cysteine-rich region
 [KW] All Alpha
 [KW] LOW_COMPLEXITY 4.90 %
 [KW] COILED_COIL 6.95 %

SEQ MPLTPTVQGFQWTLRGPVETSPFGAPRAASHGVGRHQELRDPTVPGPTSSATNVSMVVS
 SEG
 PRD ccc
 COILS

SEQ AGPWSGEKAEMNILEINKKSRPQLAENKQFRNLKQKCLVTQVAYFLANRQNNYDYEDCK
 SEG
 PRD cccccchhhhhhhheeeccccchhhhhhhhhccccchhhhhhhhhccccccccch
 COILS

[illegible]

Prosites for DKFZphtes3 7d17.2

PS000001	54->58	ASN_GLYCOSYLATION	PDOC000001
PS000001	315->319	ASN_GLYCOSYLATION	PDOC000001
PS000005	13->16	PKC_PHOSPHO_SITE	PDOC000005
PS000005	329->332	PKC_PHOSPHO_SITE	PDOC000005
PS000005	365->368	PKC_PHOSPHO_SITE	PDOC000005
PS000005	401->404	PKC_PHOSPHO_SITE	PDOC000005
PS000006	188->192	CK2_PHOSPHO_SITE	PDOC000006
PS000006	259->263	CK2_PHOSPHO_SITE	PDOC000006
PS000006	286->290	CK2_PHOSPHO_SITE	PDOC000006
PS000006	295->299	CK2_PHOSPHO_SITE	PDOC000006
PS000006	300->304	CK2_PHOSPHO_SITE	PDOC000006
PS000006	317->321	CK2_PHOSPHO_SITE	PDOC000006
PS000006	336->340	CK2_PHOSPHO_SITE	PDOC000006
PS000006	345->349	CK2_PHOSPHO_SITE	PDOC000006
PS000006	372->376	CK2_PHOSPHO_SITE	PDOC000006
PS000006	427->431	CK2_PHOSPHO_SITE	PDOC000006
PS000006	447->451	CK2_PHOSPHO_SITE	PDOC000006
PS000006	505->509	CK2_PHOSPHO_SITE	PDOC000006
PS000006	522->526	CK2_PHOSPHO_SITE	PDOC000006
PS000006	597->601	CK2_PHOSPHO_SITE	PDOC000006
PS000008	25->31	MYRISTYL	PDOC000008
PS000008	207->213	MYRISTYL	PDOC000008

Pfam for DKF2phtes3_7d17.2

HMM_NAME	TNFR/NGFR cysteine-rich region		
HMM	*CpeGtYtDWNHvpqClpCtrCePEMGQYmvqPCTwTQNTVC*		
	C +	++ +	N ++ ++ ++ ++ ++VC
Query	274	CESNQPYG-NT-RITFEEDQVDS--TLIDSSSHDEWLDAVC	310

DKFZphtes3_7j3

group: cell cycle

DKFZphtes3_7j3.2 encodes a novel 628 amino acid putative protein kinase, which is related to the C-TAK1-Cdc25C associated protein kinase.

Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to serine/threonine-specific protein kinases

complete cDNA, complete cds, potential start at Bp 128, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3443 bp

Poly A stretch at pos. 3399, polyadenylation signal at pos. 3376

```

1 GTGCTTTACT GCGCGCTCTG GTACTGCTGT GGCTCCCCGT CCTGGTGCGG
51 GACCTGTGCG CGCGGCTTCA GCCCTCCCCG CACAGCCTAC TGATCCCCCT
101 GCCGCCCCCTG CTCACCTCCT GCTCGCCATG GAGTCGCTGG TTTTCGCGCG
151 GCGCTCCGCG CCCACTCCCT CGGCCGCGAG GCTAGCCCCG CCGCTGGCGG
201 AAGGGCTGAT CAAGTCGCCC AAGCCCCCTAA TGAAGAAGCA GCGGGTGAAG
251 CGGCACCACC ACAAGCACAA CCTGCGGCAC CGCTACGAGT TCCTGGAGAC
301 CCTGGGCAAA GGCACCTACG GGAAGGTGAA GAAGGCGCGG GAGAGCTCGG
351 GCGCGCTGGT GGCCATCAAG TCAATCCGGA AGGACAAAAT CAAAGATGAG
401 CAAGATCTGA TGCACATACG GAGGGAGATT GAGATCATGT CATCACTCAA
451 CCACCCCTCAC ATCATTGCCA TCCATGAAGT GTTTGAGAAG AGCAGCAAGA
501 TCGTGATCGT CATGGAGTAT GCCAGCCGGG GCGACCTTTA TGACTACATC
551 AGCGAGCGGC AGCAGCTCAG TGAGCGCGAA GCTAGGCATT TCTTCCGGCA
601 GATCGTCTCT GCCGTGCACT ATTGCCATCA GAACAGAGTT GTCCACCGAG
651 ATCTCAAGCT GGAGAACATC CTCTTGGATG CCAATGGGAA TATCAAGATT
701 GCTGACTTCG GCCTCTCCAA CCTCTACCAT CAAGGCAAGT TCCTGCAGAC
751 ATTCTGTGGG AGCCCCCTCT ATGCCTCGCC AGAGATTGTC AATGGGAAGC
801 CCTACACAGG CCCAGAGGTG GACAGCTGGT CCCTGGGTGT TCTCCTCTAC
851 ATCCTGGTGC ATGGCACCAT GCCCTTTGAT GGGCATGACC ATAAGATCCT
901 AGTGAAACAG ATCAGCAACG GGGCCTACCG GGAGCCACCT AAACCCCTCTG
951 ATGCCTGTGG CCTGATCCGG TGGCTGTTGA TGGTGAACCC CACCCGCGCG
1001 GCCACCCCTGG AGGATGTGGC CAGTCACTGG TGGGTCAACT GGGGCTACGC
1051 CACCCGAGTG GGAGAGCAGG AGGCTCCGCA TGAGGGTGGG CACCCTGGCA
1101 GTGACTCTGC CGCGCCTCC ATGGCTGACT GGCTCCGCGG TTCCTCCCGC
1151 CCCCTCCTGG AGAATGGGGC CAAGGTGTGC AGCTTCTTCA AGCAGCATGC
1201 ACCTGGTGGG GGAAGCACCA CCCCTGGCCT GGAGCGCCAG CATTCGCTCA
1251 AGAAGTCCCG CRAAGGAGAT GACATGGCCC AGTCTCTCCA CAGTGACACG
1301 GCTGATGACA CTGCCCATCG CCCTGGCAAG AGCAACCTCA AGCTGCCAAA
1351 GGGCATTCTC AAGAAGAAGG TGTAGCCTC TGCAGAAGGG GTACAGGAGG
1401 ACCCTCCGGA GCTCAGCCCA ATCCCTGCGA GCCCAGGGCA GGCTGCCCCG
1451 CTGCTCCCCA AGAAGGGCAT TCTCAAGAAG CCCCAGCAGC GCGAGTCTGG
1501 CTACTACTCC TCTCCCGAGC CCAGTGAATC TGGGGAGCTC TTGGACGCAG
1551 GCGACGTGTT TGTGAGTGGG GATCCCAAGG AGCAGAAAGC TCCGCAAGCT
1601 TCAGGGCTGC TCCTCCATCG CAAAGGCATC CTCAAACTCA ATGGCAAGTT
1651 CTCCCAGACA GCCTTGGAGC TCGCGGCCCC CACCACCTTC GGCTCCCTGG
1701 ATGAACTCGC CCCACCTCGC CCCCTGGCCC GGGCCAGCCG ACCCTCAGGG
1751 GCTGTGAGCG AGGACAGCAT CCTGTCTCTT GAGTCTTTG ACCAGCTGGA
1801 CTTGCCTGAA CGGCTCCAG AGCCCCCACT GCGGGGCTGT GTGTCTGTGG
1851 ACAACCTCAC GGGGCTTGAG GAGCCCCCTC CAGAGGGCCC TGGAAGCTGC
1901 CTGAGGCGCT GCGGCGAGGA TCCTTTGGGG GACAGCTGCT TTTCCCTGAC
1951 AGACTGCCAG GAGGTGACAG CGACCTACCG ACAGGCACTG AGGGTCTGCT
2001 CAAAGCTCAC CTGAGTGGAG TAGGCATTGC CCCAGCCCGG TCAGGCTCTC
2051 AGATGCAGCT GGTTCACCCC CGAGGGGAGA TGCCTTCTCC CCCACCTCCC
2101 AGGACCTGCA TCCCAGCTCA GAAGGCTGAG AGGGTTTGCA GTGGAGCCCT
2151 GAGCAGGGCT GGATATGGGA AGTAGGCAAA TGAATGCGC CAAGGGTTCA
2201 GTGTCTGTCT TCAGCCCTGC TGAACGAAGA GGATACTAAA GAGAGGGGAA
2251 CGGGAATGCC CGCGACAGAG TCCACATTGC CTGTTTCTTG TGTACATGGG
2301 GGGGCCACAG AGACCTGGAA AGAGAACTCT CCCAGGGCCC ATCTCCTGCA
2351 TCCCATGAAT ACTCTGTACA CATGGTGCCT TCTAAGGACA GCTCCTTCCC
2401 TACTCATTCC CTGCCCAGT GGGGCCAGAC CTCTTTACAC ACACATTCCC
2451 TTCTCTACCA ACCACCAGAA CTGGATGGTG GCACCCCTAA TGTGCATGAG
2501 GCATCCTGGG AATGGTCTGG AGTAACGCTT CGTTATTTTT ATTTTTATTT

```

```

2551 TTATTTATTT ATTTATTTT TTGAGACGGA GTTTCGCTCT TGGTGCCAG
2601 GCTAGAGTGC AATGGCGCGA TCTCAGCTCA CCTCAACCTC CGCCTCCCGG
2651 GTTCAAGCGA TTCTCCTGCC TCAGCCTCCC TAGTAGCTGG GATTACAGGC
2701 GCGCGCCACC ATGCGCGGCT AATTTGTAT TTTTAGTAGA GACAGGGTTT
2751 CTCCATGTTG GTCAGGCTGG TCTCAAATC CCGACCTCAG GTGATCCACC
2801 CACCTCGGCC TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACCGCGCCC
2851 CACCTAACCC TTCCTTATTT AGCCTAGGAG TAAGAGAACA CAATCTCTGT
2901 TTCTTCAATG GTTCTCTTCC CTTTCCATC CTCCAAACCT GGCCTGAGCC
2951 TCCTGAAGTT GCTGCTGTGA ATCTGAAAGA CTGAAAAGC CTCCGCTGC
3001 TGTGTGGACT TCATCTCAAG GGGCCAGCC TCCTCTGGAC TCCACCTTGG
3051 ACCTCAGTGA CTCAGAACTT CTGCCTCTAA GCTGCTCTAA AGTCCAGACT
3101 ATGGATGTGT TCTCTAGGCC TTCAGGACTC TAGAATGTCC ATATTTATTT
3151 TTATGTTTCTT GGCCTTGTGT TTTAGGAAAA GTGAATCTTG CTGTTTTCAA
3201 TAATGTGAAT GCTATGTTCT GGGAAAATCC ACTATGACAT CTAAGTTTGT
3251 TGTACAGAGA GATATTTTGT CAACTATTTC CACCTCCTCC CACAACCCCC
3301 CACACTCCAC TCCCACTCTT TGAGTCTCTT TACCTAATGG TCTCTACCTA
3351 ATGGACCTCC GTGGCCAAAA AGTACCATTA AAACCAGAAA GGTGATTGGA
3401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA

```

BLAST Results

No BLAST result

Medline entries

98202387:

C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3 protein binding.

Peptide information for frame 2

ORF from 128 bp to 2011 bp; peptide length: 628
Category: strong similarity to known protein

```

1 MESLVFARRS GPTPSAAELA RPLAELIKS PKPLMKQAV KRHHKHNLR
51 HRYEFLETIG KGTYGKVKKA RESSGRLVAI KSIRKDIKD EQOLMHIRRE
101 IEIMSSSLNHP HIIAIEHVEF NSSKIVIME YASRGDLYDY ISERQQLSER
151 EARHFFRQIV SAVHYCHQNR VVHRDLKLEN ILLDANGNIK IADFGLSNLY
201 HQGKFLQTFE GSPLYASPEI VNGKPYTGPE VDSWSLGVLL YILVHGTMPF
251 DGHDKILVK QISNGAYREP PKPSDACGLI RWLLMVNPTN RATLEDVASH
301 WVVNWGYATR VGEQAPHEG GHPGSDSARA SMADWLRRSS RPLENGAKV
351 CSFFKQHAPG GGSTTPGLER QHSLKKSKE NDMAQSLHSD TADDTAHRPG
401 KSNLKLPGKI LKKKVSASAE GVQEDPPELS PIPASPGQAA PLLPKKGILK
451 KPRQRESGYI SSPESESSE LLDAGDVFSV GDPKEQKPPQ ASGLLLHRKG
501 ILKLNKGFQI TALELAAPT FGLDELAPP RPLARASRPS GAVSEDSILS
551 SESFDQLDLP ERLPEPPLRG CVSVDNLTGL EEPFSEPGFS CLRRWRQDPL
601 GDSCFSLTDC QEVTATYRQA LRVCSKLT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7j3, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_7j3; frame 2

Report for DKFZphtes3_7j3.2

```

[LENGTH]      628
[MW]           69612.39
[pI]           9.01
[HOMOL]        TREMBL:AB011109_1 gene: "KIAA0537"; product: "KIAA0537 protein"; Homo sapiens
mRNA for KIAA0537 protein, complete cds. 1e-152
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
Se-66
[FUNCAT]       11.01 stress response [S. cerevisiae, YDR477w] Se-66

```

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 5e-66
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR096w] 6e-54
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YLR096w] 6e-54
[FUNCAT] 30.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c] 8e-52
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 8e-52
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 9e-51
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 9e-51
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-45
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-42
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-34
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL139w CTK1 - carboxy-terminal domain] 2e-28
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-28
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 4e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 5e-24
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 5e-24
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHL007c] 6e-24
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 6e-24
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 1e-22
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YNR031c] 1e-22
[FUNCAT] 03.13 meiosis [S. cerevisiae, YDR523c] 8e-22
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 6e-21
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 6e-21
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 7e-19
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YDL159w] 3e-18
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 1e-17
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 1e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-17
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 4e-16
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w] 1e-15
[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w] 5e-15
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-12
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 8e-05
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 8e-05
[BLOCKS] BL00479C Phorbol esters / diacylglycerol binding domain proteins
[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[SCOP] dlqol_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 1e-77
[SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 4e-68
[SCOP] dikoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 2e-85
[SCOP] dikoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 1e-80
[SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 2e-76
[SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens)] 1e-69
[SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)] 1e-84
[SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom)] 1e-68
[SCOP] dlfdre_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)] 9e-85
[SCOP] dlfmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom)] 1e-69
[SCOP] dldcka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)] 1e-85
[SCOP] d2hcka3 5.1.1.2.1 (167-437) Haemopoetic cell kinase Hck [huma] 5e-66
[SCOP] dicsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 9e-47
[SCOP] dijsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 1e-75
[SCOP] dlckja_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 5e-54
[EC] 2.7.1.38 Phosphorylase kinase 1e-36
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 4e-40

[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 1e-61
 [EC] 2.7.1.117 Myosin-light-chain kinase 2e-40
 [EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 1e-61
 [EC] 2.7.1.37 Protein kinase 7e-42
 [PIRKW] phosphotransferase 6e-66
 [PIRKW] nucleus 1e-64
 [PIRKW] calcium 7e-35
 [PIRKW] duplication 1e-38
 [PIRKW] tandem repeat 4e-39
 [PIRKW] phorbol ester binding 1e-38
 [PIRKW] zinc 1e-38
 [PIRKW] cell cycle control 1e-42
 [PIRKW] serine/threonine-specific protein kinase 8e-68
 [PIRKW] oncogene 1e-40
 [PIRKW] phospholipid binding 1e-38
 [PIRKW] autophosphorylation 1e-64
 [PIRKW] brain 1e-40
 [PIRKW] heterotetramer 2e-36
 [PIRKW] mitosis 7e-42
 [PIRKW] polymer 1e-35
 [PIRKW] magnesium 6e-66
 [PIRKW] ATP 8e-68
 [PIRKW] polyprotein 1e-40
 [PIRKW] phosphoprotein 1e-64
 [PIRKW] apoptosis 4e-39
 [PIRKW] glycoprotein 7e-42
 [PIRKW] leucine zipper 3e-35
 [PIRKW] skeletal muscle 7e-35
 [PIRKW] protein kinase 5e-41
 [PIRKW] cAMP binding 3e-38
 [PIRKW] testis 9e-36
 [PIRKW] purine nucleotide binding 2e-49
 [PIRKW] calcium binding 8e-39
 [PIRKW] alternative splicing 3e-37
 [PIRKW] P-loop 2e-49
 [PIRKW] lipoprotein 2e-33
 [PIRKW] segmentation 1e-33
 [PIRKW] core protein 1e-40
 [PIRKW] muscle 7e-35
 [PIRKW] myristylation 2e-33
 [PIRKW] EF hand 8e-39
 [PIRKW] cell division 2e-40
 [PIRKW] calmodulin binding 4e-40
 [SUPFAM] ribosomal protein S6 kinase II 5e-36
 [SUPFAM] fibronectin type III repeat homology 3e-33
 [SUPFAM] immunoglobulin homology 3e-33
 [SUPFAM] calcium-dependent protein kinase 8e-39
 [SUPFAM] AMP-activated protein kinase 6e-66
 [SUPFAM] protein kinase akt 3e-42
 [SUPFAM] protein kinase SPK1 1e-42
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-68
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-37
 [SUPFAM] calmodulin repeat homology 8e-39
 [SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 6e-33
 [SUPFAM] protein kinase C zeta 1e-36
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-34
 [SUPFAM] death-associated protein kinase 4e-39
 [SUPFAM] pleckstrin repeat homology 3e-42
 [SUPFAM] ankyrin repeat homology 4e-39
 [SUPFAM] protein kinase homology 8e-68
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-41
 [SUPFAM] protein kinase C zinc-binding repeat homology 1e-38
 [SUPFAM] twitchin 3e-33
 [SUPFAM] protein kinase C delta 1e-38
 [SUPFAM] cGMP-dependent protein kinase 6e-33
 [SUPFAM] protein kinase cdrl 7e-42
 [SUPFAM] protein kinase C C2 region homology 3e-37
 [SUPFAM] protein kinase C alpha 3e-37
 [SUPFAM] yeast protein kinase C 5e-36
 [SUPFAM] kinase-related transforming protein 1e-41
 [SUPFAM] kinase interaction domain homology 1e-42
 [SUPFAM] gag-akt polyprotein 1e-40
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 4e-40
 [SUPFAM] protein kinase C mu 4e-33
 [PROSITE] PROTEIN_KINASE_ATP 2
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 4
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 12

```

[PROSITE]      ASN_GLYCOSYLATION      2
[PROSITE]      PROTEIN_KINASE_ST      1
[PFAM]         Eukaryotic protein kinase domain
[KW]           All_Alpha
[KW]           3D
[KW]           LOW_COMPLEXITY      10.51 %

SEQ      MESLVFARRSGPTPSAAELARPLAEGLIKSPKPLMKKQAVKRHHHKKHNLRRHYEFLETLG
SEG      .....XXXXXXXXXXXXX.....
lctpe    .....HHHHHHHHHHHHHHCCCCCCCC--GGGEEEEEEEE

SEQ      KGTYGKVKKARESSGRLVAIKSIRKDKIKDEQDLMHIRREIEIMSSLNHPHIIAIEHVFE
SEG      .....
lctpe    CTTTEEEEEETTTTEEEEEEEHHHHHHHCCCHHHHHHHHHHHHHCCCTTTBCCEEEEEEE

SEQ      NSSKIVVMYASRGDLYDYISERQQLSEREARHFFRQIVSAVHYCHQNRVVHRDLKLEN
SEG      .....
lctpe    ETTEEEEEECTTTTBHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHCCCECCCCGGG

SEQ      ILLDANGNIKIADFGLSNLYHQGKFLQTFCGSPLYASPEIVNGKPYTGPEVDSWSLGVLL
SEG      .....
lctpe    EEETTTTCEEECTTTTET-TTT-BCCCCCGGGCCHHHHHCCBC-HHHHHHHHHHHH

SEQ      YILVHGTMFPDGHDKILVKQISNGAYREPFPKPSDACGLIRWLLMVNPTRRATLEDVASH
SEG      .....
lctpe    HHHHHCCTTTTTTTHHHHHHHHHCCCTTTCHHHHHHHHHHTTTTGGGTTTHHHHHHC

SEQ      WWVNWGYATRVGEQAPHEGGHPGSDSARASMADWLRRSSRPLENGAKVCSFFKQHAPG
SEG      .....
lctpe    GG.....

SEQ      GGSTTPGLERQHSLKSRKENDMAQSLHSDTADDTAHRPGKSNLKLKPKGILKKKVSASAE
SEG      .....
lctpe    .....

SEQ      GVQEDPPPELSPIPASPGQAAPLLPKKGILKKPRQRESGYSSPEPSESSELLDAGDVFS
SEG      .....XXXXXXXXXXXXX...XXXXXXXXXXXXX.....
lctpe    .....

SEQ      GDPKEQKPPQASGLLLHRKGILKNGKFSQTALELAAPTTFGSLDELAPPRPLARASRPS
SEG      .....XXXXXXXXXXXXX.....
lctpe    .....

SEQ      GAVSEDSILSSEFQDLDLPERLPEPPLRGCVSDONLTGLEEPPSEGGPGSCLRRWRQDPL
SEG      .....XXXXXXXXXXXXX.....
lctpe    .....

SEQ      GDSCFSLTDCQEVATATYRQALRVCSKLT
SEG      .....
lctpe    .....

```

Prosites for DKFZphtes3_7j3.2

PS00001	121->125	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00004	290->294	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	337->341	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	413->417	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	142->145	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	289->292	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	373->376	PKC_PHOSPHO_SITE	PDOC00005
PS00005	377->380	PKC_PHOSPHO_SITE	PDOC00005
PS00005	616->619	PKC_PHOSPHO_SITE	PDOC00005
PS00006	15->19	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	391->395	CK2_PHOSPHO_SITE	PDOC00006

PS00006	461->465	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	578->582	CK2_PHOSPHO_SITE	PDOC00006
PS00006	606->610	CK2_PHOSPHO_SITE	PDOC00006
PS00007	453->460	TYR_PHOSPHO_SITE	PDOC00007
PS00007	453->461	TYR_PHOSPHO_SITE	PDOC00007
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	360->366	MYRISTYL	PDOC00008
PS00016	134->137	RGD	PDOC00016
PS00107	59->82	PROTEIN_KINASE_ATP	PDOC00100
PS00107	59->86	PROTEIN_KINASE_ATP	PDOC00100
PS00108	171->184	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKF2phtes3_7j3.2

HMM_NAME	Eukaryotic protein kinase domain		
HMM	*YeigRiIGeGsFGtVYkCiWrtGeIVAIIkkrrsms.....FlREI		
		YE++++G+G++G+V+K+++ +G++VAIK I+K++++ ++REI	
Query	53	YEFLETLGKGTYGKVKKARESSGRLVAIKSIRKDKIKDEQDLMHIRREI	101
HMM	qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw		
		+IM +LNHP+II + ++FE ++ I ++MEY+ GDL+DYI+++ ++SE+	
Query	102	EIMSSLNHPHIIAIEHVFE-NSSKIVIVMEYASRGDLYDYISERQQLSER	150
HMM	eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKIcDFGLARqM		
		E+R++++QI++++ Y+H ++++HRDLK ENIL+D NG+IKI+DFGL+ ++	
Query	151	EARHFFRQIVSAVHYCHQNRVVHRDLKLENILLDANGNIKIADFGLSNLY	200
HMM	nnYerMttfCGTPWYMMAPEVIImg.nyYttkVDMWSFGCILWEMMTGep		
		+ + ++ TFCG+P Y +PE+ ++G +Y +++VD WS+G++L++++ G+	
Query	201	HQGKFLQTFCGSPLYA-SPEI-VNGKPYTGPEVDSWSLGVLLYILVHGTM	248
HMM	PFyddnMemImrIiqrfrfpWpnCSeElyDFMrwCWnyDPekRPTFrQI		
		PF+++ ++ I + +++ +P S+ + ++RW++ ++P++R T +++	
Query	249	PFQGDHDKILVKQISNGAYREPPKPSD-ACGLIRWLLMVNPTRRATLEDV	297
HMM	LnHPWF*		
		H W+	
Query	298	ASHWWV 303	

DKFZphtes3_7j8

group: testes derived

DKFZphtes3_7j8 encodes a novel 410 amino acid protein nearly identical to human
WUGSC:H_DJ1159004.1.

The novel protein contains an additional C-terminal domain, which is not present in
WUGSC:H_DJ1159004.1.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific
genes.

WUGSC:H_DJ1159004.1 similarity to YBL104p

verifies and extends the genmodel WUGSC:H_DJ1159004.1
similarity to *S.cerevisiae* YBL104p

Sequenced by BMFZ

Locus: /map="7p21-p22"

Insert length: 3353 bp

Poly A stretch at pos. 3231, no polyadenylation signal found

```

1  GCAAAATATG TTGTATTGT GGCATAGTTC ATATTTACAC TATCATAAAA
51  TTATGGCCGA GAAGTTAAAT ATTCTAAATG TGTCACATA GTTCTCTGTA
101 AAACGTGACTT ATTTCCAAA TATATTTTGA AATAAAACAA TATAAAATG
151 TTTCTGTGTT TTAGGAATGG TGGAAAGCAG CAGACATAAT TGGAGTGGGT
201 TGGATAAGCA AAGTGATATT CAAAATTAA ATGAAGAGAG AATCTTAGCT
251 TTACAGCTTT GTGGGTGGAT AAAGAAAGGA ACGGATGTAG ACGTGGGGCC
301 ATTTTGAAC TCCCTTGAC AAGAAGGGGA ATGGGAAAGA GCTGCTGCTG
351 TGGCATTGTT CAACCTGGAT ATTCGCCGAG CAATCCAAAT CCTGAATGAA
401 GGGGCATCTT CTGAAAAGG AGATCTGAAT CTCAATGTGG TAGCAATGGC
451 TTTATCGGGT TATACGGATG AGAAGAACTC CCTTGGAGA GAAATGTGTA
501 GCACACTGCG ATTACAGCTA AATAACCCGT ATTGTGTGT CATGTTTGCA
551 TTTCTGACAA GTGAAACAGG ATCTTACGAT GGAGTTTGT ATGAAAACAA
601 AGTTGCAGTA CGTGACAGAG TGGCATTGTC TTGTAAATTC CTTAGTGATA
651 CTCAGTTAAA TAGATACATC GAAAAGTTGA CCAATGAAAT GAAAGAGGCT
701 GGAATTTGG AAGGAATTTT GCTTACAGGC CTTACTAAAG ATGGAGTGGA
751 CTTAATGGAG AGTTATGTTG ATAGAACTGG AGATGTTCAA ACAGCAAGTT
801 ACTGTATGTT ACAGGGTTCA CCTTTAGATG TTCTTAAAGA TGAAAGGGTT
851 CAGTACTGGA TTGAGAATTA TAGAAATTTA TTAGATGCCT GGAGGTTTGT
901 GCATAAACGA GCTGAATTTG ATATTCACAG GAGTAAAGTT GATCCCAAGT
951 CCAAGCCTTT AGCACAAGTT TTTGTGAGTT GCAATTTCTG TGGCAAGTCA
1001 ATCTCCTACA GCTGTTTCAG TGTGCCCTAT CAGGGCAGAG GTTTTAGTCA
1051 GTATGGTGTG AGTGGCTCAC CAACGAAATC TAAAGTCACA ATTTGTCCTG
1101 GCTGTCGAAA ACCACTTCCT CGATGTGCGC TTTGTCTCAT TAATATGGGA
1151 ACACCAGTTT CTAGCTGTCC TGGAGGAACC AAATCAGATG AAAAAGTGGG
1201 CTTGAGCAAG GACAAAAAAT TAGCCCAATT TAACAAGTGG TTTACATGGT
1251 GTCAATATTG CAGGCACGGT GGACATGCTG GACATATGCT TAGTTGGTTC
1301 AGGGACCATG CAGAGTGCCC TGTGTCTGCA TGCACGTGTA AATGTATGCA
1351 GTTGGATACA ACGGGGAATC TGGTACCTGC AGAGACTGTC CAGCCATAAA
1401 ATGTTACCAC CTTAAGAGAA CCCTTCAAGT GTGGAGCTTT CTAGTAGGTG
1451 TCCTTCATAG CTCAGAAACA TACCTCAGAA CAAGCCATTC ATGACTTACC
1501 TGTAAATGGG AATAAATCA TTCTATCAGA TCAGCAGTTT TGATGTTTGA
1551 GTGATTTTGA TATGCTTCAC AGAGACAAAT GCTGCCAAAA TAAACATCGA
1601 AGTATAGACA TGAGTTCTGT TCAGCAGGTT GAAAAGTCTG ATTTAGAAAA
1651 ACTTTCTAAG TTTTGGTTGA AATTATGAAC ACTCTAGAAG CAGAATTTCT
1701 GGAAGAGCCA AGAAGAGACT TTGAGCCTAT ATCTTCAAAG CTGAAACTGG
1751 ATATCTTTCA ATAAAATAG TGCACTTTAA AAATAAAATG ACTAATTTCT
1801 TGATTCAGAC AATAGTTTAA AGTTCAGCTG TGCTTAGATT TCTTTCAGAT
1851 TAATTTAAAA TTATAGATTT TACTTTTAG AATTGCAGAG CCCCTATCCC
1901 ACACCTGGGA ATATTTTTTA TTAGTGTCTG TTATATATGT GTCTATGTGT
1951 GTGTGTATAT TTATGTGTGT ATGTATAAAT ATGTACTTTT TAAAGGAGCC
2001 TTTTCCCTCC TTTGATTTTA AGATAAGCAA TCTTTTGCCA TAACATTATC
2051 GTCTTCCTAG AAAAGCCAAG ATGAAGAAATC TATCTTACAA CTTTTCCTCT
2101 TCACTAGAGA AAAACATGTA CCATTTCAGG TGAACATACA AAATTTTCAC
2151 TTTTACCTTT TTGCCTTCCA ATGTCCTGAT TTGTCTTCAA AGGTTTTTCT
2201 CCATATTAAAT TTGTATCTT ATCCTCATCA CCTGAGAAAC TTTTACTGCA
2251 TACAAGTCT ATGCAAGATT ATATGTAAC AGCCATTTAG TATAATCTAT
2301 GTCAGTGTGT CTGTGCTGTC AAATTCCTGC CTGATTGGA ATACCATACC
2351 TTGTTCTTTC CAAGGTAGAC TAGGAAGTGT TGGGGAAATA GGGTCACTTC
2401 AGAGACCATT TTAGATGTAA GTTTTTAAAT GTAAGTGTTA CTGGGGCTAA
2451 GTCAGGGACT TTATTTAAAA CATTTTTTTT TTCTCATTTT ATAGCTAGAT
2501 AGTTGTAAGA GAAATACAAA GAATTTACAA GATGCTTCTC TGTCATCTGC

```

```

2551 CGTATGCAGA GGGACTGAAC TAGGAATTTT GTAGTTGAAG CTGTGTTTCAT
2601 AAAGAGTAAA TCTTATTTTA TAGATTTTGG AGAAATAAA CAAGAATTTT
2651 AAGAGCTTTC GTATTAGCAG TTTTGCCTTA TAAAACTAA GATTGTGCAG
2701 ATTAGTTTGA GGTGTAACCT AAATATTTAA AGTAGATTAA ATTTATTTT
2751 TACCTTGAGT GTCTGATACA TAAACCCCTT TTCTAGGAAA ACATTGGAAG
2801 TAGTACATAT TACTCTAAA TGTCTCACCT GCATGACAGT CTTTCAAAT
2851 GAAAGACATG GTAATTGCAA TTTTTTTTAA AAGATTGCTA TTAAGGGTAC
2901 TTTTCCAGC CTTCAATTGA GTAAATCTTA ATTGATTTC ATTTATTAAC
2951 ATATACCCCTT TACCTTTAAT ATTTTCATTG AAGTGTTCCT TTCAAACCTA
3001 CTGTCTTAAA TATGAAAGTC AGCTTTAAGT AATGTCAGAC TCATATGCAT
3051 TTTCAATTCTC ATTAGCTAAA GTAAATGTGA AAATATCTC AAATAGTTAC
3101 AAGTTTTGGA AATACAGTAT AAAACATGAA TGTAAAGTCT ATTATGTAAT
3151 ATGCTTATTT GTAATCCTAA TATATGAGGG TGACATTTT AAGATTGTAT
3201 GTATGTGTCA ACCTCTTAAA TGTTTTCTGT GAAAAAAAAA AAAAAAAAAA
3251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 167 bp to 1396 bp; peptide length: 410
 Category: known protein
 Classification: unclassified

```

1 MVESSRHNWS GLDKQSDIQN LNEERILALQ LCGWIKKGTD VDVGPFLNSL
51 VQEGEWERAA AVALFNLDIR RAIQILNEGA SSEKGDNLN VVAMALSGYT
101 DEKNSLWREM CSTLRLQLNN PYLCVMFAFL TSETGSYDGV LYENKVAVRD
151 RVAFACKFLS DTQLNRYIEK LTNEMKEAGN LEGILLTGLT KDGVDLMESY
201 VDRGTGDVQTA SYCMLQGSPL DVLKDERVQY WIENYRNLLD AWREWHKRAE
251 FDIHRSKLDP SSKPLAQVFV SCNFCGKSIS YSCSAVPHQG RGSFYQGVSG
301 SPTKSKVTSC PGCCKPLPRC ALCLINMGTP VSSCPGGTKS DEKVDLSKDK
351 KLAQFNWFT WCHNCRHGGH AGHMLSWFRD HAECPVSACT CKCMQLDTTG
401 NLVPAETVQP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzptes3_7j8, frame 2

PIR:S45391 probable membrane protein YBL104c - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 446, P = 4.5e-47

TREMBL:AC004982.1 gene: "WUGSC:H_DJ1159004.1"; Homo sapiens PAC clone
 DJ1159004 from 7p21-p22, complete sequence., N = 1, Score = 2038, P =
 7.6e-211

>TREMBL:AC004982.1 gene: "WUGSC:H_DJ1159004.1"; Homo sapiens PAC clone
 DJ1159004 from 7p21-p22, complete sequence.
 Length = 379

HSPs:

Score = 2038 (305.8 bits), Expect = 7.6e-211, P = 7.6e-211
 Identities = 379/379 (100%), Positives = 379/379 (100%)

```

Query:      1 MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDDVDVGPFLNSLVQEGEWERAA 60
            MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDDVDVGPFLNSLVQEGEWERAA
Sbjct:      1 MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDDVDVGPFLNSLVQEGEWERAA 60

Query:     61 AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN 120
            AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN
Sbjct:     61 AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN 120

```


Query:	121	PYLCVMFAFLTSETGSDGVLYENKAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN	180
		PYLCVMFAFLTSETGSDGVLYENKAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN	
Sbjct:	121	PYLCVMFAFLTSETGSDGVLYENKAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN	180
Query:	181	LEGILLTGLTKDGVDMLESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD	240
		LEGILLTGLTKDGVDMLESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD	
Sbjct:	181	LEGILLTGLTKDGVDMLESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD	240
Query:	241	AWREWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSIYSYCSAVPHQGRGFSQYGVSG	300
		AWREWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSIYSYCSAVPHQGRGFSQYGVSG	
Sbjct:	241	AWREWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSIYSYCSAVPHQGRGFSQYGVSG	300
Query:	301	SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT	360
		SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT	
Sbjct:	301	SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT	360
Query:	361	WCHNCRHGGHAGHMLSFR	379
		WCHNCRHGGHAGHMLSFR	
Sbjct:	361	WCHNCRHGGHAGHMLSFR	379

Pedant information for DKFZphtes3_7j8, frame 2

Report for DKFZphtes3_7j8.2

```
[LENGTH]      410
[MW]           45862.45
[pI]           6.51
[HOMOL]        TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159004.1"; Homo sapiens PAC clone DJ1159004
from 7p21-p22, complete sequence. 0.0
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YBL104c] 7e-48
[BLOCKS]       BL00028 Zinc finger, C2H2 type, domain proteins
[BLOCKS]       BL00534A Ferrochelatase proteins
[PIRKW]        transmembrane protein 2e-46
[KW]           All Alpha
```

[illegible]

(No Prosite data available for DKFZphtes3_7j8.2)

(No Pfam data available for DKFZphtes3_7j8.2)

DKF2phtes3_7p10

group: Cell Cycle

DKF2phtes3_7p10.1 encodes a novel 422 amino acid putative protein, which is closely related to the *Xenopus laevis* XPMC2 protein.

In fission yeast the kinases Wee1 and Mik1 control that initiation of mitosis starts after completion of DNA synthesis. Yeast in which both Wee1 and Mik1 kinases are defective exhibit a mitotic catastrophe phenotype. XPMC2 of *xenopus* rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function. The XPMC2 protein is localised in the nucleus in *Xenopus* oocytes. The new protein is the human orthologue of this gene.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to XPMC2 protein

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="9q34"

Insert length: 2380 bp

Poly A stretch at pos. 2341, polyadenylation signal at pos. 2318

```
1 AGCGTGCGTG CTGAGGTATG CGCAACGCGT GCGGGGTCTC TTCCGGAGTC
51 TTTTCTTGGG CCGGGTCCCT GCGGTGGGTG TGTTTCGGCC TGGCCTGGGC
101 AGGCGCTTGT GCTGCCAGGG CGCCGGGCCC GGGGAGGCCG GGGTCTCGGG
151 TGGCCGCGCG CCCAGGCCTG GGACGGCAGC AGGATGGGGA AGGCGAAGGT
201 CCCCCTCTCC AAGCGCGCCC CGAGCAGCCC CGTGGCTAAG CCGGGTCTTG
251 TCAAGACGCT CACTCGGAAG AAAAACAAGA AGAAAAAAG GTTTTGGAAA
301 AGCAAGGCGC GGGAAAGTAA CAAGAAGCCA GCAAGCGGCC CCGGTGCTGT
351 GGTGCGACCT CCAAAGGCAC CAGAAGACTT TTCTCAAAAC TGGAAAGGCG
401 TGCAAGAGTG GCTGCTGAAA CAAAATCTC AGGCCCCAGA AAAGCCTCTT
451 GTCATCTCTC AGATGGGTTC CAAAAGAAG CCCAAATTA TCCAGCAAAA
501 CAAAAAAGAG ACCTCGCCTC AAGTGAAGGG AGAGGAGATG CCGGCAGGAA
551 AAGACCAGGA GGCCAGCAGG GGCTCTGTTC CTTAGGTTTC CAAGATGGAC
601 AGGAGGGGCG CAGTACCTCG CACCAAGGCC AGTGAACAG AGCACAATAA
651 GAAAGGAACC AAGGAAAGGA CAAATGGTGA TATTGTTCCT GAACGAGGGG
701 ACATCGAGCA TAAGAAGCGG AAAGCTAAGG AGGCAGCCCC AGCCCCACCC
751 ACCGAGGAAG ACATCTGGTT TGACGACGTG GACCCAGCGG ATATCGAAGC
801 TGCCATAGGT CCAGAGGCGG CCAAGATAGC GAGGAAACAG TTGGGTGAGA
851 GCGAGGGGCG CGTCAGCCTC AGCCTCGTGA AAGAGCAGGC CTTGCGGCGG
901 CTGACAAGAG CCTTAGCCTT GGACTGTGAG ATGGTGGGCG TGGGCCCTAA
951 GGGGGAGGAG AGCATGGCCG CCCGTGTGTC CATCGTGAAC CAGTATGGGA
1001 AGTGCCTTTA TGACAAGTAC GTCAAACCAA CTGAGCCCGT GACGGACTAT
1051 AGGACAGCGG TCAGTGGGAT TCGGCCCTGAG AACCTCAAGC AGGGAGAAGA
1101 GCTTGAAGTT GTTCAGAAGG AAGTGGCAGA GATGCTGAAG GGCAGAATTC
1151 TAGTGGGGCA CGCTCTGCAT AATGACCTAA AGGTACTATT TCTTGATCAT
1201 CCAAAAAAGA AGATTGCGGA CACACAGAAA TATAAACCTT TCAAGAGTCA
1251 AGTAAAGAGT GGAAGGCCGT CTCTGAGACT ACTTTCAGAG AAGATCCTTG
1301 GGCTCCAGGT CCAGCAGGCG GAGCACTGTT CAATTCAGGA TGCCAGGCA
1351 GCAATGAGGC TGTACGTCAT GGTGAAGAAG GAGTGGGAGA GCATGGCCCCG
1401 AGACAGGCGC CCCCTGCTGA CTGCTCCAGA CCACTGCAGT GACGACGCCCT
1451 AGCAGTCTCG CCCTGCTGCT GCTGCCGCCG CGCTACAGAG GCAATGTGAC
1501 CAGTCACAGG GACAGATCAC ATCTCCCCAG AGTGCAACT CTGGTGAAC
1551 CTTTTACAGT TCATGGCAGA GGGGCGTGGC GTGGTGCTAC TGAGAAGGTC
1601 CTCCTTCCTC TTGACTTTGT GGTCTGAAAC CTGGTCTTAC TGTCCATGTG
1651 TGTTTGGGCC CGGATGGTCA GGGTGGGGAG CAGGGACGGC CATGGGCACG
1701 CCTGGCCACG CTTTACCGAC TGCTGACCCC CTGGGCCAGG TGAGGTTGGG
1751 GCCTGTGGGC CGCCAGTCCA TACGGTGCTG TCACTGCCCA TCTTCGGTGA
1801 CACCTGGGGT TGAGGTGCTC AGCACCTTCC TCTCGAGGAG CCACATTTTC
1851 CTCCTTTGTG TTAGGGGACA TAACAAGCTC TGCTGGGCTT GAGGGACCCA
1901 GACCAAGGTG CTGCAGTCAG CTCCTGAGAC ACAGCTGGCC GGCACAACAG
1951 GTGTTACATC AGGGGTTTCC TGTGGCCGTT TGAACTTTGA GCATTTATCT
2001 AAATTAAATT GGCCAGGGGT TGGCTGGTGG GTCAACCAGC AGAGGCTTCT
2051 CCCCATAGCA CGAGGATGTG TTGCCCTGGC ACGGTGACTG CGGTATTATC
2101 TGGAGGTCCG CAGACATGCC AACCTTGGGC TATTTGAGCT GGAGAAGCTA
2151 TGTGATGCTA GCCCGTGGCT TTCTGGGCTA GGCCCCAGTT TGAGGCTCCC
2201 CTGGGAAGTA GAGCCAGGAA CAGCCAGTGG CACTGACAAG GGGACGGAGT
2251 CCAAGGCGTT ATTGGGCCAC CTGACAGCTG GACAGAAAAG GGGCAGACAC
2301 ACCGAGGATG CGATTTAAAA TAAATGCAGA TGTTTACTTG GAAAAAATAA
2351 AAAAAAATAA AAAAAAATAA AAAAAAATAA
```

BLAST Results

Medline entries

Peptide information for frame 1

1	MGKAKVPASK	RAPSSPAKAP	GPVKTLTRKK	NKKKKRFWKS	KAREVSKKPA
51	SGPGAVVRPP	KAPEDVFSQNW	KALQEWLLQG	KSQAPKPLVR	ISQMGSKKKP
101	KIIQKQKKET	SPQVKGSEMP	KAGQDEASRG	SVPSGSKMDR	RAPVPTKRSK
151	GTEHNKKGTK	ERTKGDIQVE	RGDIEHKKRK	AKEAAPAPPT	EEDIWFDQVD
201	PADEAAATGP	EAAKATARKLQ	QSGSVSVLS	LVKEAPFGLG	TRALALDCGM
251	VGVGPKGEES	MAARVSVINQ	YGKCVQDKYV	LPTEQVTDYR	TAVSGIRPEN
301	LKQGEELVV	QKEVAEMLKG	RILVGHALHN	DLKVLFDLHP	KKKTRDTQKY
351	KPFKFSQVSG	RPSRLRLSEK	ILGLQVQAAE	HCSIQDAQAA	MRLYVMVKKE
401	WESMARDRRP	LLTAPDHCSD	DA		

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7p10, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_7p10, frame 1

Report for DKFZphtes3 7p10.1

```

[LENGTH]      422
[MW]           46671.91
[pI]           9.79
[HOMOL]        PIR:S53818 XPMC2 protein - African clawed frog 7e-96
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-42
[FUNCAT]       01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 2e-19
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 7e-13
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YGL094c] 7e-13
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YLR107w] 6e-10
[PROSITE]      RGD 1
[PROSITE]      MYRISTYL 4
[PROSITE]      CAMP_PHOSPHO_SITE 2
[PROSITE]      CK2_PHOSPHO_SITE 6
[PROSITE]      TYR_PHOSPHO_SITE 2
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 8
[KW]           All Alpha
[KW]           LOW COMPLEXITY 11.37 %

```

```

SEQ      MGKAKVPASRRAPSSPVAKPGPVKTLTRKKNKKKKRFRWFSKAREVSKKPSAGSPGAVVRFP
SEG      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      RAPEDFSQNWKALQEWLLKQKSQAPEKPLVISQMGSKKKPKIIQONKRETS PQVKGEMP
SEG      .....xxxxxxxxxxxxx.....
PRD      cccccccchhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccce

```

```

SEQ      AGKDQAEASRGSVPSGSKMDRRAPVPRTKASGTEHNKKGTKERTNGDIVPERGDIHEKKRK
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      AKEAAPAPPTTEEDIWFDDVPADI EAAIGPEAAKIARKQLGQSEGSVLSLSVKEQAFGG
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhccccccccceeeccccchhhhhhhccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhh

SEQ      TRALALDCEMVGVPKGEESMAARVSIVNQYKGCYDYKVPKTEPVTDYRTAVSGIRPEN
SEG      .....
PRD      hhhccccccccccccchhhhhhhhhccccccccceeeeeecccccccccccccccccccccc

SEQ      LKQGEELEVQKEVAEMLKGRILVGHALHNDLKVFLDHPKKKIRDTQKYKPFKSQVKS
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhccceeeccccchhhhhhhhhhhccccccccceeecccccccccc

SEQ      RPSLRLLSEKILGLQVQQAHEHCSIQDAQAAMRLYVMVKKEWESMARDRRPLLTAPDHCS
SEG      .....
PRD      chhhhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc

SEQ      DA
SEG      ..
PRD      CC

```

Prosites for DKFZphtes3_7p10.1

PS000002	51->55	GLYCOSAMINOGLYCAN	PDOC00002
PS000004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	156->160	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS000005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS000005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS000005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS000005	347->350	PKC_PHOSPHO_SITE	PDOC00005
PS000005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS000005	363->366	PKC_PHOSPHO_SITE	PDOC00005
PS000005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS000006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS000006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS000006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS000006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS000006	383->387	CK2_PHOSPHO_SITE	PDOC00006
PS000006	413->417	CK2_PHOSPHO_SITE	PDOC00006
PS000007	343->351	TYR_PHOSPHO_SITE	PDOC00007
PS000007	342->351	TYR_PHOSPHO_SITE	PDOC00007
PS000008	130->136	MYRISTYL	PDOC00008
PS000008	151->157	MYRISTYL	PDOC00008
PS000008	221->227	MYRISTYL	PDOC00008
PS000008	239->245	MYRISTYL	PDOC00008
PS000016	171->174	RGD	PDOC00016

(No Pfam data available for DKFZphtes3_7p10.1)

DKFZphtes3_7p9

group: nucleic acid management

DKFZphtes3_7p9 encodes a novel 691 amino acid protein with similarity to human nuclear domain 10 protein NDP52.

The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle.

The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex.

The new protein can find application in modulation of viral infections and tumour events.

similarity to nuclear domain 10 protein NDP52

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="329.1 cR from top of Chr12 linkage group"

Insert length: 3003 bp

Poly A stretch at pos. 2957, no polyadenylation signal found

```

1 AAGGTGAGGG GAACAGCTGA TCCGTCTGTT GGGAGGACAG ATATCTCAAG
51 GCCAGGATGG AAGAATCACC ACTAAGCCGG GCACCATCCC GTGGTGGAGT
101 CAACCTTTCTC AATGTAGCCC GGACCTACAT CCCCACACAC AAGGTGGAAT
151 GTCACCTACAC CCTTCCCCCA GGCACCATGC CCAGTGCCAG TGAAGTGGAT
201 GGCATCTTCA AGGTGGAGGC TGCCTGTGTT CGGGATTACC ACACATTGTT
251 GTGGTCTTCC GTGCCTGAAA GTACAACTGA TGGTTCCCCC ATTCACACCA
301 GTGTCCAGTT CCAAGCCAGC TACCTGCCCA AACCAGGAGC TCAGCTCTAC
351 CAGTTCCGAT ATGTGAACCG CCAGGGCCAG GTGTGTGGGC AGAGCCCCCC
401 TTTCCAGTTT CGAGAGCCAA GGCCCATGGA TGAAGTGGTG ACCCTGGAGG
451 AGGCTGATGG GGGCTCTGAC ATCCTGCTGG TTGTCCCAA GGCAGTGTG
501 TTACAGAACC AGCTCGATGA GAGCCAGCAA GAACGGAATG ACCTGATGCA
551 GCTGAAGCTA CAGCTGGAGG GACAGGTGAC AGAGCTGAGG AGCCGAGTGC
601 AGGAGCTCGA AGGGGCTCTG GCAACTGCCA GGCAGGAGCA CACGGAGCTG
651 ATGCAACAGT ACAAGGGGAT TTCCCGGTCC CATGGGGAGA TCACAGAAGA
701 GAGGGACATC CTGAGCCGGC AACAGGGAGA CCATGTGGCA CGCATCCTGG
751 AGCTAGAGGA TGACATCCAG ACCATCAGTG AGAAAGTGCT GACGAAGGAA
801 GTGGAGCTGG ACAGGCTTAG AGACACAGTG AAGGCCCTGA CTCGGGAACA
851 AGAGAAGCTC CTTGGGCAAC TGAAAGAAGT ACAAGCAGAC AAGGAGCAAA
901 GTGAGGCTGA GCTCCAGTG GCACAACAGG AGAACCATCA CTTAAATTTG
951 GACCTGAAGG AGGCGAAGAG CTGGCAAGAG GAGCAGAGTG CTCAGGCTCA
1001 GCGACTGAAA GACAAGGTGG CCCAGATGAA GGACACCCTA GGCCAGGCCC
1051 AGCAGCGGGT GGCCGAGCTG GAGCCCTTGA AGGAGCAGCT TCGAGGGGGC
1101 CAGGAGCTTG CAGCCTCAAG CCAGCAGAAA GCCACCCTTC TTGGGGAGGA
1151 GTTGGCCAGC GCAGCAGCAG CCAGGGACCG CACCATAGCC GAACTACACC
1201 GCAGCCGCTT GGAAGTGGCT GAAGTTAACG GCAGGCTGGC TGAGCTCGGT
1251 TTGCACTTGA AGGAAGAAAA ATGCCAATGG AGCAAGGAGC GGGCAGGGCT
1301 GCTGCAGAGT GTGGAGGCAG AGAAGGACAA GATCCTGAAG CTGAGTGCAG
1351 AGATACTTCG ATTGGAGAAG GCAGTTCAGG AGGAGAGGAC CCAAAACCAA
1401 GTGTTCAAGA CTGAGCTGGC CCGGGAGAAG GATTCTAGCC TGGTACAGTT
1451 GTCAGAAAGT AAGCGGGAGC TGACAGAGCT GCGGTACGCC CTGCGTGTGC
1501 TCCAGAAGGA AAAGGAGCAG TTACAGGAGG AGAAACAGGA ATTGCTAGAG
1551 TACATGAGAA AGCTAGAGGC CCGCCTGGAG AAGGTGGCAG ATGAGAAGTG
1601 GAATGAGGAT GCCACCACAG AGGATGAGGA GGCCGCTGTG GGGCTGAGCT
1651 GCCCGGCAGC TCTGACAGAC TCAGAGGACG AGTCCCCAGA AGACATGAGG
1701 CTCCCACCCT ATGGCCTTTG TGAGCGTGGG GACCCAGGCT CCTCTCCTGC
1751 TGGGCTCTGA GAGGCTTCTC CCCTTGTGTT CATCAGCCAG CCGGCTCCCA
1801 TTTCTCCTCA CCTCTCTGGG CCAGCTGAGG ACAGTAGCTC TGACTCGGAG
1851 GCTGAAGATG AGAAGTCAGT CCTGATGGCA GCTGTGCAGA GTGGGGGTGA
1901 GGAGGCCAAC TTAAGTCTTC CTGAAGTGGG CAGTGCCCTC TATGACATGG
1951 CCAAGTGGGT TACAGTGGGT ACCCTGTGAG AAACCCAGCA TGGGGGGCCT
2001 GCCACCCCCA CATGGAAGGA GTGTCTATC TGTAAAGGAGC GCTTTCCTGC
2051 TGAGAGTGAC AAGGATGCCC TGGAGGACCA CATGGATGGA CACTTCTTTT
2101 TCAGCACCCA GGACCCCTTC ACCTTTGAGT GATCTTACTC CCTCGTACAT
2151 GCACAAATAC ACACTCATGC ACACACACAC TCACACACAT GCATACACTT
2201 AGGTTTATC CTCCATTTCT ATCACACTGG GCTCCATGAT ATTCTGTTCC
2251 CTAAGAACTG CTTCTGTGTG CCCTGTTTTT ATCCCAAGAT TTCTCACTTC
2301 ATCCTCTCCT ACCTGGCTCT TTTGTCCAG GAGGGGGTCC TGTTTCGGAAG
2351 CAGTGGCTGA ATTTATCCCC TGAAAGTGGT TTTGGAGGAA CCGGGATGGA
2401 GGAGGCCCTC CCCTGTGGGA ATAGAATCGT CCACTCCTAG CCCTGGTTGC

```

```

2451 TTCTGATACA CAGCCACTGC ACACACACAC TCACACTCAC ACTCCCTTGT
2501 CTGATGCCCC AAAGCCAATT CCTGGGGCAC CCTACCTCTT CTTATTGGGA
2551 GTTTCGGTTG GTTTACCTGA GTTTTCTCTG GGGTCTGCAC AGAGGCAGCA
2601 GCATGGACAT CATGGCCTCT CAGGTCCCTT TTGGTCTCA GTTTCATTGG
2651 TTCCTCTTTC TGTTCCTCCA TTGACTCTGT TGCCCCACCC TAGCCTTTTC
2701 CATAACCTTA GGTATTTCAGT TTGGAGGGGT TTTTGTATT TTTGAGGATT
2751 CCTGTATTCT GTATCCTCTC CTCGCATCTC CTCACATGGA AAGAAATAAT
2801 GTATTGTGTC CTTCTGTGAG GAATGGGGGG AACAAAGTGT CCCAGGTATC
2851 CCCATTTCCT AGGCCCCCTT CCCTCTCCAG GTCCCCCAC AGCAATAAAA
2901 GCTTCCCCTT GATATCCATC CCTTTGTAGT TTGAACAAAT ATATTTATAT
2951 GATATGTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3001 AAA

```

BLAST Results

Entry HS189353 from database EMBL:
human STS WI-11261.
Score = 2191, P = 1.4e-92, identities = 463/485

Medline entries

95310349:
Molecular characterization of NDP52, a novel protein of the nuclear domain 10, which is redistributed upon virus infection and interferon treatment.

97375672:
Cellular localization, expression, and structure of the nuclear dot protein 52.

Peptide information for frame 3

ORF from 57 bp to 2129 bp; peptide length: 691
Category: similarity to known protein
Prosites motifs: RGD (557-560)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (475-497)
LEUCINE_ZIPPER (482-504)

```

1 MEESPLSRAP SRGGVNFLNV ARTYIPNTKV ECHYTLPPGT MPSASDWIGI
51 FKVEAACVRD YHTFVWSSVP ESTTDGSPH TSVQFOASYL PKPGAQLYQF
101 RYVNRQGVQC GQSPFFQFRE PRPMDELVTI EEADGSDIL LVVPKATVLQ
151 NQLDESQQR NDLMLKLQL EGVTELRSR VQELERALAT ARQEHTELME
201 QYKGISRSHG EITEERDILS RQQGDHVAR LELEDDIQT SEKVLTKEVE
251 LDRLRDTPKA LTREQKLLG QLKEVQADKE QSEAEQVAQ QENHHLNLDL
301 KEAKSWQEEQ SAQAQRLKDK VAQMKDTLGG AQQRVAELEP LKEQLRGAQE
351 LAASSQKAT LLGEELASAA AARDRTIAEL HRSRLEVAEV NGRLAELGLH
401 LKEEKCQWSK ERAGLLQSVE AEKDKILKLS AEILRLEKAV QEERTONQVF
451 KTELAREKDS SLVQLSESKR ELTELRSALE VLQKEKEQLQ EEKQELLEYM
501 RKLEARLEKV ADEKWNEDAT TEDEEAAVGL SCPAALTDSE DESPEDMRP
551 PYGLCERGGP GSSPAGPREA SPLVVISQPA PISPHLSGPA EDSSSDSEAE
601 DEKSVLMAAV QSGGEEANLL LPELGSAFYD MASGFTVGTL SETSTGGPAT
651 PTWKECPICK ERFPAESDKD ALEDHMDGHF FFSTQDPFTF E

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7p9, frame 3

PIR:A56733 nuclear domain 10 protein NDP52 - human, N = 2, Score = 307,
P = 7.7e-28

TREMBL:AB008852_1 gene: "NDP"; product: "NDP52"; Bos taurus mRNA for NDP52, complete cds., N = 2, Score = 302, P = 4e-27

TREMBL:AC004549_1 gene: "WUGSC:H RG459N13.1"; product: "TXBP151"; Homo sapiens BAC clone RG459N13 from 7p15, complete sequence., N = 2, Score = 275, P = 2.3e-25

PIR:G02043 TXBP151 - human, N = 2, Score = 270, P = 8.5e-25

TREMBL:DM35816.4 gene: "zip"; product: "nonmuscle myosin-II heavy chain"; Drosophila melanogaster nonmuscle myosin-II heavy chain (zip) gene, complete cds., N = 1, Score = 254, P = 1.4e-17

>PIR:A56733 nuclear domain 10 protein NDP52 - human
Length = 446

HSPs:

Score = 307 (46.1 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
Identities = 104/323 (32%), Positives = 158/323 (48%)

```
Query: 15 VNFLNVARTYIPNTKVECHYTLPPTGTMPSASDWIGIFKVEAACVRDYHTFVWSSVPESTT 74
      V F +V + YIP V CHYT +P DWIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFNSVEKFYIPGGDVTCHYTFTQHFIPRRKDWIGIFRVGWKTTREYTYTFMWVTLPIDLN 82

Query: 75 DGSPHITSVQFQASYLKPKGAQLYQFRYVNRQGVCGQSPFPQFREPRPMDELVTLEED 134
      + S VQF+A YLPK + YQF YV+ G V G S PFQFR D LV +
Sbjct: 83 NKSAAQQEVQFKAYYLPKDD-EYYQFCYVDEGCVVRGASIPFQFRPENEEDILVVTQ-- 139

Query: 135 GGS DILLVVPKATVLQNLQ-LDES---QOERNLMLQLQLLEGQVTE-LRSRVQELERALA 189
      G + + K +NQ L +S Q++N MQ +LQ + + E L+S ++LE +
Sbjct: 140 GEVEEIEQHKNELCKENQELKDCISLQKQNSDMQAELOKKQEELETLSINKKLELKV 199

Query: 190 TARQE-HTELMEQYKGISRSHGEITEERDI-LSRQGDHVARILELEDDIQTISEKVLTK 247
      + TEL+ Q K ++ E+ I + + Q + E+E +Q +K T+
Sbjct: 200 EQKDYWETELL-QLKEQNQKMSSENEKMGIRVDQLQAQLSTQEKEMKLVQGDQDK--TE 256

Query: 248 EVE-LDRLRDTVKALTREQEKLGLQKEVQADKEQSEAELOVAQQENHHLNLDLKEAKSW 306
      ++E L + D + EQ K +L++ +Q+E QQE N DL + S
Sbjct: 257 QLEQLKKENDHFLSLTEQRKDQKKLEQTVQMKQNETTAMKKQQLMDENFDLSKRLSE 316

Query: 307 QEEQSAQAQRLKDKVAQMKTDLGQAQQRV 335
      E QR K+++ D L + R+
Sbjct: 317 NEIICNALQRQKERLEGENDLLKRENSRL 345
```

Score = 304 (45.6 bits), Expect = 2.1e-27, Sum P(2) = 2.1e-27
Identities = 98/337 (29%), Positives = 163/337 (48%)

```
Query: 15 VNFLNVARTYIPNTKVECHYTLPPTGTMPSASDWIGIFKVEAACVRDYHTFVWSSVPESTT 74
      V F +V + YIP V CHYT +P DWIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFNSVEKFYIPGGDVTCHYTFTQHFIPRRKDWIGIFRVGWKTTREYTYTFMWVTLPIDLN 82

Query: 75 DGSPHITSVQFQASYLKPKGAQLYQFRYVNRQGVCGQSPFPQFREPRPMDELVTLEED 134
      + S VQF+A YLPK + YQF YV+ G V G S PFQFR P +E
Sbjct: 83 NKSAAQQEVQFKAYYLPKDD-EYYQFCYVDEGCVVRGASIPFQFR---PENE----- 130

Query: 135 GGS DILLVVPKATVLQNLQDESQOERNLMLQLQLLEGQVTELSRVQELERALATARQE 194
      DIL+V Q +++E +Q +L + +L+ L+ + +++ L +QE
Sbjct: 131 --EDILVVT-----QGEVEEIEQHKNELCKENQELKDCISLQKQNSDMQAELOK-KQE 182

Query: 195 HTELMQYKGISRSHGEITEERDILSRQGDH-VARILELEDDIQTISEKVLTKVELDR 253
      E ++ I ++ ++ ++Q D+ +L+L++ Q +S + + +D+
Sbjct: 183 ELETLS-----INKKLELKVKEQKDYWETELLQLKEQNQKMSSENEKMGIRVDQ 232

Query: 254 LRDTVKALTREQEKL--GOLKEVQAD---KEQSEAELOVAQQENHHLNLDLKEAKSWQE 308
      L+ + +E EKL+ Q K Q + KE L + +Q L+ + Q
Sbjct: 233 LQAQLSTQEKEMKLVQGDQDKTEQLEQLKKENDHFLSLTEQRKDQKKLEQTVQMKQN 292

Query: 309 EQSA--QAQRLKDKVAQMKTDLGQAQQRVAELEPLKEQLRGAQL 351
      E +A + Q L D+ + L + + L+ KE+L G +L
Sbjct: 293 ETTAMKKQQLMDENFDLSKRLSENEIICNALQRQKERLEGENDL 337
```

Score = 124 (18.6 bits), Expect = 2.3e-06, Sum P(2) = 2.3e-06
Identities = 53/227 (23%), Positives = 113/227 (49%)

```
Query: 138 DILLVVPKATVLQNLQDESQOERNLMLQLQLLEGQVTELSRVQELERALATARQEHTE 197
      DIL+V Q +++E +Q +L + +L+ L+ + +++ L +QE E
Sbjct: 132 DILVVT-----QGEVEEIEQHKNELCKENQELKDCISLQKQNSDMQAELOK-KQEELE 185

Query: 198 LMEQYKGISRSHGEITEERDILSRQGDH-VARILELEDDIQTISEKVLTKVELDR 256
      ++ I ++ ++ ++Q D+ +L+L++ Q +S + + +D+L+
Sbjct: 186 TLQS-----INKKLELKVKEQKDYWETELLQLKEQNQKMSSENEKMGIRVDQLQA 235

Query: 257 TVKALTREQEKLGLQKEVQADKEQSEAELOVAQQENHHLNLDLKEAKSWQEEQSAQAR 316
      + +E EKL VQ D+++E +L+ ++EN HL L L E + Q++ ++
Sbjct: 236 QLSTQEKEMKLVQGDQDKTE-QLEQLKKENDHFLSLTEQRKDQKKLEQTVQ 288
```

Query: 317 LK-DKVAQMKDTLGQAQQRVAELEPLKEQLRGAQELA-ASSQKATLLGE 364
 +K ++ MK + Q+ + E L ++L ++ A +QK L GE
 Sbjct: 289 MKQNETTAMK---KQELMDENFDLSKRLSENEIICNALQRQKERLEGE 334

Score = 103 (15.5 bits), Expect = 4.4e-04, Sum P(2) = 4.4e-04
 Identities = 63/278 (22%), Positives = 123/278 (44%)

Query: 299 DLKEAKSWQEEQSAQAQRLKDKVAQMK---DTLGQAQQRVAELEPLKEQLRGAQELAAS 354
 +++E + +E + Q LKD ++ D + Q++ ELE L + + EL
 Sbjct: 141 EVVEIEQHNLCKENQELKQKSCISLQKQNSDMQAELOKKQEELETLSINKKLELKV 199

Query: 355 SQKATLLGEELASAAAARDRTIAELHRSRLEVAEVNGRLAELGLHLKEEKCQWSKERAG 414
 Q+ EL + +E + + V ++ +L+ + E+ Q +++
 Sbjct: 200 EQKD--YWETELLQKEQKQKMSSENEKMGIRVDQLQAQLSTQEKEM-EKLVQGDQDKTE 256

Query: 415 LLQSVAEAKDKI-LKLSAEIL---RLEKAVQEERTONQVFKTELAREKDSLSVLSESKE 470
 L+ ++ E D + L L+ + +LE+ V E+ QN+ T + +++ SKR
 Sbjct: 257 QLEQLKKENDHLFLSLTEQRKQKLEQTV-EQMKQNET--TAMKKQELMDENFDLSK 313

Query: 471 ELTELRSALRVLQKEKEQLQEEKQELLEYMRKLEARLEKVADEKWNE---DATTEDEEAA 527
 L+E LQ++KE+L+ E +LL ++ +RL +N T DE A
 Sbjct: 314 -LSENEIICNALQRQKERLEGEN-DLL---KRENSRLSYMGDFNSLPYQVPTSDEGGA 368

Query: 528 ---VGLSCPAALTD-SEDESPEDMRLLPPYGLCERGDGSSPAGPREASPL 573
 GL+ + E SP ++ +C+ D ++ PL
 Sbjct: 369 RQNPGLAYGNPYSGIQESSSPSPLSIKKCPICKADDICDHTLEQQQMQL 418

Score = 64 (9.6 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
 Identities = 13/29 (44%), Positives = 17/29 (58%)

Query: 651 PTWKECPICKERFFAESDKDALEDHMDGH 679
 P CPIC + FPA ++K EDH+ H
 Sbjct: 417 PLCFNCPICDKIFPA-TEKQIFEDHVFCH 444

Score = 64 (9.6 bits), Expect = 5.8e+00, Sum P(2) = 1.0e+00
 Identities = 26/90 (28%), Positives = 45/90 (50%)

Query: 470 RELTELRSALRVLQKEKEQLQEE---KQELLEYMRKLEARLE-KVADEK--W----- 515
 +E EL+ + LQK+ +Q E QKE LE ++ + +LE KV ++K W
 Sbjct: 154 KNEQELKQKSCISLQKQNSDMQAELOKKQEELETLSINKKLELKVKEQKDYWETELLQLK 213

Query: 516 --NEDATTEDEEAAVGLS-CPAALTDSEDE 542
 N+ ++E+E+ + + A L+ E E
 Sbjct: 214 EQNQKMSSENEKMGIRVDQLQAQLSTQKE 243

Score = 47 (7.1 bits), Expect = 4.6e-26, Sum P(2) = 4.6e-26
 Identities = 11/30 (36%), Positives = 17/30 (56%)

Query: 631 MASGFTVGTLSSETSTGGPATPTWKECPICK 660
 +A G + E+S+ P + K+CPICK
 Sbjct: 374 LAYGNPYSGIQESSSPSPLSI--KKCPICK 401

Pedant information for DKFZphtes3_7p9, frame 3

Report for DKFZphtes3_7p9.3

[LENGTH] 691
 [MW] 77336.52
 [pI] 4.77
 [HOMOL] PIR:A56733 nuclear domain 10 protein NDP52 - human 2e-29
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 2e-11
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 2e-11
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 2e-11
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 2e-11
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 2e-11
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 2e-08
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YJL074c] 4e-07
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL250w] 4e-06
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YBR289w] 4e-06

[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YBR289w] 4e-06

[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YBR289w] 4e-06

[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 4e-06

[FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 4e-06

[FUNCAT] 1 genome replication, transcription, recombination and repair (M. jannaschii, MJ1643] 1e-05

[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 4e-05

[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-05

[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 7e-05

[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 7e-05

[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 7e-05

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c] 2e-04

[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-04

[BLOCKS] BL00682B 2P domain proteins

[EC] 3.6.1.32 Myosin ATPase 1e-13

[PIRKW] nucleus 6e-10

[PIRKW] phosphotransferase 2e-07

[PIRKW] duplication 9e-07

[PIRKW] citrulline 1e-09

[PIRKW] tandem repeat 1e-13

[PIRKW] heart 5e-11

[PIRKW] endocytosis 5e-09

[PIRKW] polymorphism 3e-06

[PIRKW] cornified cell envelope 1e-06

[PIRKW] transmembrane protein 6e-12

[PIRKW] serine/threonine-specific protein kinase 2e-07

[PIRKW] cell wall 1e-06

[PIRKW] zinc finger 5e-09

[PIRKW] metal binding 5e-09

[PIRKW] DNA binding 8e-08

[PIRKW] muscle contraction 1e-11

[PIRKW] IgG constant region-binding 1e-06

[PIRKW] acetylated amino end 4e-09

[PIRKW] actin binding 1e-13

[PIRKW] mitosis 9e-09

[PIRKW] microtubule binding 9e-09

[PIRKW] ATP 1e-13

[PIRKW] thick filament 1e-10

[PIRKW] phosphoprotein 1e-13

[PIRKW] epidermis 1e-06

[PIRKW] leucine zipper 1e-07

[PIRKW] glycoprotein 4e-07

[PIRKW] skeletal muscle 4e-10

[PIRKW] disulfide bond 1e-07

[PIRKW] calcium binding 1e-09

[PIRKW] alternative splicing 1e-10

[PIRKW] coiled coil 1e-13

[PIRKW] P-loop 1e-13

[PIRKW] heptad repeat 6e-10

[PIRKW] methylated amino acid 1e-13

[PIRKW] basement membrane 3e-06

[PIRKW] immunoglobulin receptor 2e-07

[PIRKW] peripheral membrane protein 5e-09

[PIRKW] dimer 1e-07

[PIRKW] cardiac muscle 1e-10

[PIRKW] extracellular matrix 3e-06

[PIRKW] hydrolase 1e-13

[PIRKW] microtubule 6e-10

[PIRKW] muscle 2e-09

[PIRKW] membrane protein 3e-06

[PIRKW] EF hand 1e-09

[PIRKW] cytoskeleton 6e-12

[PIRKW] hair 1e-09

[PIRKW] calmodulin binding 5e-09

[PIRKW] Golgi apparatus 3e-08

[SUPFAM] myosin heavy chain 1e-13

[SUPFAM] conserved hypothetical P115 protein 1e-08

[SUPFAM] hypothetical protein YJL074c 5e-07

[SUPFAM] centromere protein E 9e-09

[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-07

[SUPFAM] calmodulin repeat homology 1e-09

[SUPFAM] myosin motor domain homology 1e-13

[SUPFAM] alpha-actinin actin-binding domain homology 3e-13

[SUPFAM] tropomyosin 3e-07

[SUPFAM] plectin 3e-13

[SUPFAM] trichohyalin 1e-09

[SUPFAM] pleckstrin repeat homology 4e-06

[SUPFAM] ribosomal protein S10 homology 3e-13

[SUPFAM] giantin 3e-08
 [SUPFAM] protein kinase homology 2e-07
 [SUPFAM] protein kinase C zinc-binding repeat homology 4e-06
 [SUPFAM] involucrin 1e-06
 [SUPFAM] kinesin motor domain homology 9e-09
 [SUPFAM] human early endosome antigen 1 5e-09
 [SUPFAM] unassigned kinesin-related proteins 8e-08
 [SUPFAM] M5 protein 3e-08
 [SUPFAM] cytoskeletal keratin 3e-08
 [PROSITE] LEUCINE_ZIPPER 3
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 6
 [PROSITE] CK2_PHOSPHO_SITE 25
 [PROSITE] PKC_PHOSPHO_SITE 6
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 9.12 %
 [KW] COILED_COIL 39.36 %

SEQ MEESPLSRAPSRGGVFNFLNVARTYIPNTKVECHYTLPPGTMPSASDWIGIFKVEAACVRD
 SEG
 PRD ccc
 COILS

SEQ YHTFWSSVPESTTDGSPIHSTVQFQASYLPKPGAQLYQFRVNRQGVCGQSPFFQFRE
 SEG
 PRD eeeeeeeccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccccc
 COILS

SEQ PRPMDLVLEADGGSDILLVVPKATVLQNQLDESQQERNDLMQLKLQLEGQVTELRSR
 SEG
 PRD cccccceehhhhhchhh
 COILScc

SEQ VQELERALATARQEHTELMEQYKGISRSHGEITEERDILSRQQGDHVARILELEDDIQT
 SEG
 PRD hhh
 COILS ccc

SEQ SEKVLTKVELDRLRDTVKALTRQEKLGLQKEVQADKEQSEAEQVAQQENHHLNLDL
 SEG
 PRD hhh
 COILScc

SEQ KEAKSWQEEQSAQAQRLKDKVAQMMDTLGQAQQRAVELEPLKEQLRGAQELAASSQQRAT
 SEG
 PRD hhh
 COILS ccccc..cc

SEQ LLGEELASAAAARDRTIAELHRSRLEVAEVNRLAELGLHLKEKQWQSKERAGLLQSVE
 SEG xxx
 PRD hhh
 COILS cccccccc.....cccccccccccc

SEQ AEKDKILKLSAEILRLKAVQEERTQNQVFKTELAREKDSSLVQLSESKRELTELRSALR
 SEG
 PRD hhh
 COILS ccc

SEQ VLQKEKEQLQEEQOELEMYMRKLEARLEKVADEKWNEDATTEDEAAVGLSCPAALTDSE
 SEG .xx
 PRD hhh
 COILS ccc

SEQ DESPEDMRLPPYGLCERGDGSSPAGPREASPLVVISQPAPISPHLSGPAEDSSSDSEAE
 SEG
 PRD hhhccchh
 COILS

SEQ DEKSVLMAAVQSGGEEANLLPELGSAFYDMASGFTVGTLSETSTGGPATPTWKECPICK
 SEG xx.....
 PRD hhhhhhhhhhhhhcc
 COILS

SEQ ERFPAESDKDALEDHMDGHFFSTQDPFTFE
 SEG
 PRD cccccccchhhhhhhcccccecccccccc
 COILS

Prosites for DKFZphtes3_7p9.3

PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	468->471	PKC_PHOSPHO_SITE	PDOC00005
PS00005	652->655	PKC_PHOSPHO_SITE	PDOC00005
PS00005	667->670	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDOC00006
PS00006	376->380	CK2_PHOSPHO_SITE	PDOC00006
PS00006	383->387	CK2_PHOSPHO_SITE	PDOC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	537->541	CK2_PHOSPHO_SITE	PDOC00006
PS00006	539->543	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00006	595->599	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00006	612->616	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	652->656	CK2_PHOSPHO_SITE	PDOC00006
PS00006	667->671	CK2_PHOSPHO_SITE	PDOC00006
PS00006	683->687	CK2_PHOSPHO_SITE	PDOC00006
PS00008	39->45	MYRISTYL	PDOC00008
PS00008	107->113	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	414->420	MYRISTYL	PDOC00008
PS00008	561->567	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00016	557->560	RGD	PDOC00016
PS00029	163->185	LEUCINE_ZIPPER	PDOC00029
PS00029	475->497	LEUCINE_ZIPPER	PDOC00029
PS00029	482->504	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_7p9.3)

DKFZphtes3_8e24

group: signal transduction

DKFZphtes3_8e24.3 encodes a novel 658 amino acid putative GTP-binding protein, related to yeast YGL099w and mouse MMR1 putative GTP-binding proteins.

GTP-binding proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

strong similarity to guanine nucleotide binding proteins

complete cDNA, complete cds, potential start at Bp 31, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 3290 bp

Poly A stretch at pos. 3269, polyadenylation signal at pos. 3251

```
1 CGTCCAGCGG TCGTGTGGCC ATGGGCCCGGA GGAGAGCCCC GGCCGGTGGG
51 TCGCTGGGAC GGGCCCTTAT GCGCCATCAG ACTCAGCGGA GCCGAAGCCA
101 TCGTCACACT GACTCCTGGT TGCACACAAG TGAACCAAT GATGGCTATG
151 ATTGGGGTCG TCTTAATCTT CAGTCAGTGA CTGAACAGAG CTCCCTTGAT
201 GACTTCCTTG CTAAGTGCAG ACTTGCAGGA ACAGAGTTTG TAGCTGAAAA
251 ACTTAATATT AAGTTTGTGC CTGCTGAGGC TAGAAGTGA CTAAGTCTTT
301 TCGAGGAGAG CCAGAGAATT AAGAAGCTCC ATGAAGAAAA CAAACAGTTC
351 TTGTGTATAC CGAGGAGACC AAACTGGAAC CAAATACTA CCCCAGAAGA
401 ACTCAAAACA GCAGAGAAAG ATAACCTTCT AGAATGGAGA CGTCAGCTTG
451 TCCGGCTAGA AGAGGAACAG AAGCTGATAT TGAACGAAAT TGAACGAAAT
501 TTGGACTTTT GCGGCCAGCT CTGGAGAGTC ATTGAGAGAA GTGATATTGT
551 GGTCCAGATA GTAGATGCTC GAAACCCACT CCTGTTTAGA TGTGAGGATT
601 TGGAAATGTT TGTGAAAGAA ATGGATGCCA ATAAGGAGAA CGTCATTCTG
651 ATCAACAAGG CAGACTTGCT GACTGCTGAG CAGCGGAGTG CCTGGGCCAT
701 GACTTTCGAA AAAGAAGATG TGAAGGTTAT TTTCTGGTCA GCTTTGGCCG
751 GAGCCATTCC CTTGAATGGT GACTCTGAGG AAGAGGCAAA CAGAGATGAT
801 AGACAAAGCA ACACAACTGA GTTTGGACAT TCCAGTTTCG ACCAGGCTGA
851 AATTTCACAC AGTGAATCCG AACATCTCCC AGCTAGGGAT TCTCCTTAC
901 TTAGTGAAAA TCCCACAACG GATGAAGATG ACAGTGAGTA TGAGGACTGT
951 CCAGAGGAGG AGGAAGACGA CTGGCAGACG TGCTCAGAAG AAGACGGTCC
1001 CAAGGAAGAG GACTGCAGCC AGGACTGGAA GGAAAGCTCT ACTGCAGATT
1051 CTGAGGCTCG GAGCAGGAAA ACCCCACAGA AGAGGCAGAT ACACAATTTT
1101 AGCCATCTGG TATCCAAGCA GGAGTTACTG GAGCTCTTTA AGGAGCTACA
1151 CACTGGGAGA AAGGTGAAAG ATGGGCAACT TACGGTCGGA CTGGTGGGCT
1201 ACCCTAATGT TGGTAAGAGT TCAACAATCA ACACCATCAT GGGCAACAAG
1251 AAAGATCTCT GTCTGTCCAC ACCTGGTCAC ACAAGCACTT TTCAGACTCT
1301 CTATGTGGAG CCTGGCCTCT GCCTGTGTGA CTGTCCTGGC TTGGTGATGC
1351 CATCTTTTGT GTCTACCAAG GCAGAAATGA CTGTCAGCGG AATCCTCCCA
1401 ATTGATCAGA TGAGAGATCA TGTTCTCTCT GTATCACTAG TTTGCCAGAA
1451 TATTCCAAGA CATGTTTTAG AAGCTACCTA TGGCATTAAAC ATCATAACGC
1501 CTAGAGAGGA TGAAGATCCC CACCGACCTC CAACATCGGA AGAAGTGTG
1551 ACAGCTTATG GATACATGCG AGGATTCTAT ACAGCGCATG GACAGCCAGA
1601 CCAGCCTCGA TCTGCGCGCT ACATCCTGAA GCACTATGTC AGTGGTAAAG
1651 TGCTGTACTG CCATCCTCCT CCTGGAAGAG ATCCTGTAAC TTTTCAGCAT
1701 CAACACACAG GACTCCTAGA GAACAAAATG AACAGTGATG AAATAAAAAAT
1751 GCAGCTAGGC AGAAATAAAA AAGCAAAGCA GATTGAAAAT ATCGTTGACA
1801 AAACCTTTTT CCATCAAGAG AATGTGAGGG CTTTGACCAA AGGAGTCCAG
1851 GCTGTGATGG GTTACAAGCC CGGGAGTGGT GTAGTGACTG CATCCACTGC
1901 GAGCTCTGAG AACGGGGCGG GGAAGCCCTG GAAAAACAT GGCAACAGAA
1951 ATAAAAAAGA AAAAAATCGT AGACTCTACA AGCACCTGGA TATGTGAGGT
2001 TGGGCTGCAA CAGAAATGTC ATCTGCATTG TGCAGATGGA AAAGAGCAGA
2051 AGCTGCCTGT TGCCTGTGGA ACTGTCCCAA GACACTAGCA CTGTAGAACG
2101 GGCCCTGTCT TTGCAGAGCA CGGCTGCACC CAACAGTCTC CATGTCAAGA
2151 CCAAGGGGCT CCTGGAAACA CCAGCTCTGA CAAAAAGGAG TCATCTGGGA
2201 GCCCGAGAAT CCTACTCCTG GCCGGGCACA GTGGCTCAGC CACCAACATG
2251 GAGAAACCCC GTCTCTACTA AAAATACAAA AAAATTAGCC AGGCGTGGTG
2301 GCGCGCACCT GTAATCCCAG CTACTCGGGA GGCTGAGGCA AGGAGATCAC
2351 TTGAACCAAG GAGGCAGAGT TTGCAGTGAA TGGAGATTGC GCCGCTGCAC
2401 TCCAGCCTGG GCGACAGAGT GAGACTGCAT CACAAGAAAA AAAATTGCA
2451 AGGGATGGTT CACGAGACAC ATTTGGGACG AAGGTGAAAG AGAAATTCCT
2501 CATCTGAGT GTCCCTAGTT GGTTCCTCCG ACTCTAAACA AGGGAAGTTG
2551 GTTCAGTTAG TGTACAGCGG GGGCTCACGT CCACTAAGGA ACATGTAGAA
2601 TGTAACCAAC GGGTGACAGG GAAGCTGCGG TATTTACTAC CTAGCCCCCA
```

```

2651 TCTTCACTGG TTATTCCACT TATTTAAAT GTCCAGAATA AGCAAATCTC
2701 CATATAGAGG AAGTAGATTA GTGGTIGCTT CGGGATGGGA GGAATGGGAA
2751 GATTGAGGTC TTTCTTTTGC AGTGATAAAA ATGTCTTAAA ATTGACTGTA
2801 GCGATGGTCA CACAACTCTG AATATGCTTA AGACCATTGA ATTACACACT
2851 TTACGTTGGT GAATTGTATG GTATGTAAAT TATAGTTCAA TAACATAGTT
2901 ACAAAGATA ATCAAAGCA TGAAAGCACT ATTGATGTGG TTTGGATCTG
2951 TGTCTCTACC GAGTCTCATG TTGAAATGTA AGCCCCCTGG TGGGAGGCGA
3001 TGGGATTATG GGGCAGAGTC CTCACAAACG GTTAGCACC ACCCGCTCAG
3051 TGCTGTTCTC CTGATATTGA GTCCTCATCA CATCTGGTTG CTTCAAAGTG
3101 TGTGGTGCCT CCCCTCTGTC TCCCTCCTGC TCTGGCCATA TAAGATGTGC
3151 CTGCTTCTCC TTCGCTTCT AACATGATTG TAAGTTTCTT GAGGCCTCCC
3201 TAGAAGCAAA AGCTGCTGTG CTTCTGTAC CATCTACTGG ACCGTGAGCC
3251 AATTAAACCT CTTTCTTTA TAAAAAAGG AAAAAAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 21 bp to 1994 bp; peptide length: 658
 Category: strong similarity to known protein

```

1 MGRRRAPAGG SLGRALMRHQ TQSRSHRHT DSWLHTSELN DGYDWGRNLN
51 QSVTEQSSLD DFLATAELAG TEFVAEKLNI KFVPAEARTG LLSFEESQRI
101 KKLHEENKQF LCIPRRPNWN QNTTPEELKQ AEKDNFLEWR RQLVRLEEEQ
151 KLILTPFERN LDFWRQLWRV IERSDIVVQI VDARNPLLFR CEDLECYVKE
201 MDANKENVIL INKADLLTAE QRSAWAMYFE KEDVKVIFWS ALAGAIPLNG
251 DSEEEANRDD RQSNTEFGH SSFDQAEISH SESEHLPARD SPSLSENPTT
301 DEDDSEYEDC PEEEEDDWQT CSEEDGPKKE DCSQDWKESS TADSEARSRK
351 TPQKRQIHNF SHLVSKQELL ELFKELHTGR KVKDQQLTVG LVGYPNVGKS
401 STINTIMGNK KVSVSATPGH TKHFQTLTYVE PGLCLDCDPC LVMPSPVSTK
451 AEMTCSGILP IDQMRDHVPP VSLVCQNIPI HVLEATYGIN IITPREDEDP
501 HRPPTSEELL TAYGYMRGFM TAHGQPDQPR SARYILKDYV SGKLLYCHPP
551 PGRDPVTFQH QHQRLLENKM NSDEIKMQLG RNKKAKQIEN IVDKTFHFQE
601 NVRALTKGVQ AVMGYKPGSG VVTASTASSE NGAGKPWKKH GNRNKKKESR
651 RLYKHLDM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzptes3_8e24, frame 3

SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN
 CHROMOSOME I., N = 3, Score = 560, P = 1.6e-111

PIR:S64106 hypothetical protein YGL099w - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 544, P = 2.6e-105

TREMBL:CEAF3143_1 gene: "C53H9.2"; Caenorhabditis elegans cosmid
 C53H9., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MMR1_MOUSE POSSIBLE GTP-BINDING PROTEIN MMR1., N = 2, Score =
 311, P = 7.5e-31

>SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN
 CHROMOSOME I.
 Length = 616

HSPs:

Score = 560 (84.0 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
 Identities = 119/253 (47%), Positives = 163/253 (64%)

Query: 12 LGRALMRHQ TQSRSHRHT DSWLHTSELN DGYDWGRNLN QSVTEQSSLD DFLATAELAGT 71
 LGRA+ T+ R+ + H + + R L+SVT ++ LD+FL TAEI

Sbjct: 12 LGRAIQSDFTKNRRNRK--GGLKHIVSDPKAH--RAALRSVTHETDLDEFLNTAELGEV 67

Query: 72 EFVAEKLNIKFVP-AEARTGLLSFEESQRIKKLHEENKQFLCIPRRPNWNQNTTPEELKQ 130
EF+AEK N+ + E LLS EE+ R K+ E+NK L IPRRP+W+Q TT EL +

Sbjct: 68 EFIAEKQNVTVIQNPEQNPFLLSKEEAARSKQKQKKNKDRLTIPRRPHWDQTTAVELDR 127

Query: 131 AEKDNFLEWRRQLVRLLEEQKLILTPFERNLDFWRQLWRVIERSDIVVQIVDARNPLLF 190
E+++FL WRR L +L++ + I+TPFERNL+ WRQLWRVIERSD+VVQIVDARNPL FR

Sbjct: 128 MERESFLNWRRLAQLQDVEGFIVTPFERNLIEWRQLWRVIERSDVVVQIVDARNPLFFR 187

Query: 191 CEDLECYVKEMDANKENVILINKADLLTAEQRSAWAMYFEKEDVKVIEWSALAGAIPLNG 250
LE YVKE+ +K+N +L+NKAD+LT EQR+ W+ YF + ++ +F+SA . A N

Sbjct: 188 SAHLEQYVKEVGPSKKNFLLVNKADMLTEEQRNYWSSYFNENNIPFLFFSARMAA-EANE 246

Query: 251 DSEEEANRDDRQSN 264
E+ + SN

Sbjct: 247 RGEDLETYESTSSN 260

Score = 532 (79.8 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 131/323 (40%), Positives = 192/323 (59%)

Query: 340 STADSEARSRKTPQKRQIHNFSHLVSKQELLEFLKELHTGRKVKDQ--LTVGLVGYPNV 397
ST+ +E + +H+ S + + + L +F++ + + DG+ +T GLVGYPNV

Sbjct: 256 STSSNEIPESLQADENDVHS-SRIATLKVLEGIFEKFA--TLPDGKTKMTFGLVGYPNV 312

Query: 398 GKSSTINTIMGNKKVSVSATPGHTKHFTLYVEPGLCLDCPGLVMPFSVSTKAEMTCSG 457
GKSSTIN ++G+KKVSVS+TPG TKHFQT+ + + L DCPGLV PSF +T+A++ G

Sbjct: 313 GKSSTINALVGSKKVSSTPGTKHFQTINLSEKVSLLDCPGLVFPFATTQADLVLDG 372

Query: 458 ILPIDQMRDHVPPVSLVCQNIPIRHVLEATYGINI-ITPREDEDPHRPPTSEELLTAYGYM 516
+LPIDQ+R++ P +L+ + IP+ VLE Y I I I P E E P+++E+L +

Sbjct: 373 VLPIDQLREYTGPSALMAERIPKEVLETLYTIRIRIKPIE-EGGTGVPSAQEVLPFARS 431

Query: 517 RGFMTAH-GQPDQPRSAARYILKDYVSGKLLYCHPPPG--RDPVTFQHQHQRLLLENKMNSD 573
RGFM AH G PD R+AR +LKDYV+GKLLY HPPP F +H + + + SD

Sbjct: 432 RGFMRHHGTPODSRAARILLKDYVNGKLLYVHPPPNYPNSGSEFNKEHHQKIVSA-TSD 490

Query: 574 EIKMQLGR---NKKAKQIEN-IVDKTFFHQEN--VRALTQGVQAVM-G--YKPGSGVVTA 624
I +L R + E+ +VD +F QEN VR + KG M G YK + +

Sbjct: 491 SITEKLQRTAISDNILSAESQLVDDEYF-QENPHVRPMVKGTAAMQGPVYKGRNTMQPF 549

Query: 625 STASSENGAGK-PWKKHGNRNKKEKSRRL 652
+++ + K P G + K+R+L

Sbjct: 550 QRRNLDDASPKYPMNAQGGKPLSRKARQL 578

Score = 47 (7.1 bits), Expect = 1.3e-60, Sum P(3) = 1.3e-60
Identities = 21/84 (25%), Positives = 35/84 (41%)

Query: 552 GRDPVTFQHQHQRLLLENKMNSDEIKMQLGRNKKAKQIENIVDKTFFHQENVRALTQGVQA 611
G D T++ + + +DE + R K +E I +K F TK

Sbjct: 248 GEDLETYESTSSNEIPESLQADENDVHSSRIATLKVLEGIFEK--FASTLPDGKTKMTFG 305

Query: 612 VMGYKPGSGGVVTAASSENGAGK 635
++GY P G +ST ++ G+ K

Sbjct: 306 LVGY-PNVG--KSSTINALVGSKK 326

Score = 43 (6.5 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 7/13 (53%), Positives = 9/13 (69%)

Query: 638 KKHGNRNKKEKSR 650
KKH +NK+ K R

Sbjct: 596 KKHNNKKNRSKQR 608

Pedant information for DKFZphtes3_8e24, frame 3

Report for DKFZphtes3_8e24.3

[LENGTH] 658
[MW] 75226.58
[pI] 5.86
[HOMOL] SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME
I. 5e-56
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL099w] 3e-55
[FUNCAT] r general function prediction [M. jannaschii, MJ1464] 1e-16
[FUNCAT] 08.16 extracellular transport [S. cerevisiae, YER006w] 3e-09
[PIRKW] P-loop 1e-27
[PIRKW] GTP binding 1e-27
[SUPFAM] conserved hypothetical protein MG442 7e-08

```

[PROSITE]      ATP_GTP_A      1
[PROSITE]      MYRISTYL      3
[PROSITE]      AMIDATION      2
[PROSITE]      CAMP_PHOSPHO_SITE      1
[PROSITE]      CK2_PHOSPHO_SITE      19
[PROSITE]      TYR_PHOSPHO_SITE      2
[PROSITE]      PKC_PHOSPHO_SITE      10
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      4.56 %

SEQ  MGRRRAPAGSGSLGRALMRHQTQSRSRSHRHTDSWLHTSELNDGYDWGRNLQSVTEQSSLD
SEG  .....XXXXXXXXXXXXX.....
PRD  cccccccccchhhhhhhhhccccccccccccccccccccccccchhhhhhhhhccccch

SEQ  DFLATAELAGTEFVAEKLNIKFVPAEARTGLLSFEESQRIKKLHEENKQFLCIPRRPNWN
SEG  .....
PRD  hhhhhhhhhheeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccc

SEQ  QNTTPEELKQAEKDNFLEWRRLVRLLEEQLILTPFERNLDFWRQLWRVIERSDIVVQI
SEG  .....
PRD  cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeee

SEQ  VDARNPLLFRCEDLECYVKEMDANKENVILINKADLLTAEQRSAMWYFEKEDVKVIFWS
SEG  .....
PRD  eccccccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccc

SEQ  ALAGAIPLNGDSEEEANRRDROSNTTEFGHSSFDQAEISHSESEHLPARDSPSLSENPTT
SEG  .....
PRD  cccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ  DEDDSEYEDCPPEEEDDWQTCSEEDGPKCEDCSQDWKESSTADSEARSRKTQKQRIHNF
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccc

SEQ  SHLVSKQELLELFKELHTGRKVKDGQLTVGLVGYPNVGKSSINTIMGNKKVSVSATPGH
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

SEQ  TKHFQTLYVEPGLCLDCPGLVMPFSFVSTKAEMTCSGILPIDQMRDHPVPPVSLVCQNIPR
SEG  .....
PRD  cceeeeeeeccccccccccccccccccccchhhhhhhhhccccccccccccccccccccccccch

SEQ  HVLEATYGINIITPREDEDPHRPPTSEELLTAYGYMRGFMTHAGQPDQPRSARYILKDYV
SEG  .....
PRD  hhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhcc

SEQ  SGKLLYCHPPPPGRDPVTFQHQHQRLLKNMNSDEIKMQLGRNKKAKQIENIVDKTFFHQE
SEG  .....
PRD  cceeeeeccccccccccccchhhhhhhhhhhccccchhhhhhhhhccchhhhhhhhhhhhhhhhhccch

SEQ  NVRALTKGQVAMGYKPGSGVVTASTASSENGAGKPWKHGNRNKKEKSRRLYKHLDM
SEG  .....
PRD  hhhhhhhheeeeeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhcc

```

Prosites for DKFZphtes3_8e24.3

PS00001	264->268	ASN_GLYCOSYLATION	PDOC00001
PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00004	410->414	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00005	448->451	PKC_PHOSPHO_SITE	PDOC00005
PS00005	493->496	PKC_PHOSPHO_SITE	PDOC00005
PS00005	531->534	PKC_PHOSPHO_SITE	PDOC00005
PS00005	541->544	PKC_PHOSPHO_SITE	PDOC00005
PS00005	649->652	PKC_PHOSPHO_SITE	PDOC00005
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	155->159	CK2_PHOSPHO_SITE	PDOC00006
PS00006	252->256	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006

PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	299->303	CK2_PHOSPHO_SITE	PDOC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	505->509	CK2_PHOSPHO_SITE	PDOC00006
PS00007	480->488	TYR_PHOSPHO_SITE	PDOC00007
PS00007	190->198	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	432->438	MYRISTYL	PDOC00008
PS00008	620->626	MYRISTYL	PDOC00008
PS00009	1->5	AMIDATION	PDOC00009
PS00009	378->382	AMIDATION	PDOC00009
PS00017	393->401	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphtes3_8e24.3)

DKFZphtes3_8g11

group: testes derived

DKFZphtes3_8g11 encodes a novel proline-rich 939 amino acid protein without similarity to known proteins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop).
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin rich protein

1 EST hit (from testis library)

Sequenced by MediGenomix

Locus: unknown

Insert length: 3100 bp

Poly A stretch at pos. 3056, polyadenylation signal at pos. 3041

```
1 AGAGTCTTCC CTCAGCATAT TTTACGATAG AGAAGATCTT GTTCCAATGG
51 AAGAAAGTGA GGACTCACAG AGTGATTCCC AGACAAGGAT TTCTGAGTCC
101 CAACACTCCC TCAAGCCAAA TTATCTTTCC CAGGCCAAGA CTGACTTCTC
151 AGAACAGTTC CAGTTGCTAG AAGATCTGCA GCTAAAAATA GCAGCAAAAC
201 TCTTAAGGAG TCAAATACCC CCCGATGTGC CTCCACCTCT AGCTTCAGGT
251 CTAGTCCCTAA AATACCCTAT CTGCCTACAG TGTGGCCGAT GTTCAGGACT
301 TAATTGCCAT CATAAATTAC AGACCACTTC GGGGCCTTAT CTTCTTATCT
351 ATCCACAGCT CCACCTTGTA CGCACTCCTG AAGGCCATGG TGAGGTTCCGG
401 TTGCATCTTG GCTTTAGGCT GAGAATTGGG AAAAGATCCC AAATCTCAAA
451 GTATCGTGAA AGAGATAGAC CCGTCATACG GAGAAGCCCT ATATCACCAT
501 CACAAAGGAA AGCTAAAATC TATACTCAAG CTTCCAAGAG TCCTACTTCC
551 ACAATAGATT TGCAGTCTGG GCCTTCCCAG TCCCCTGCTC CTGTACAAGT
601 CTACATCAGG CGAGGACAAC GCAGCAGGCC TGACTTAGTA GAAAAGACAA
651 AACTAGAGC ACCTGGGCAC TATGAATTCA CTCAAGTTCA CAACCTACCA
701 GAGAGTGACT CTGAAAGCAC TCAGAATGAA AAACGGGCTA AAGTGAGAAC
751 CAAAAAGACC TCTGATTCAA AATATCCAAT GAAGAGAATC ACCAAGCGAC
801 TTAGAAAACA CAGAAAGTTC TACACAAACA GTAGAACCAC AATAGAGAGT
851 CCTTCTAGGG AATTAGCAGC CCATTTAAGA AGGAAGAGGA TTGGAGCAAC
901 TCAGACAAGT ACTGCCTCTT TAAAAAGACA ACCTAAGAAA CCTTCCCAAC
951 CCAAGTTCAT GCAACTGCCT TTTCAGAGCC TAAAGCGGGC ATTCCAAACA
1001 GCACACAGAG TTATAGCTTC TGTGGGCGG AAGCCTGTGG ACGGGACAAG
1051 GCCAGACAAT TTGTGGGCAA GCAAAACTA TTATCCAAA CAAATGCCA
1101 GGGACTATTG CTTACCAAGC AGTATCAAAA GAGACAAGAG GTCAGCTGAC
1151 AAGCTAACGC CAGCAGGCTC AACCATTAA GAGGAGGACA TATTGTGGGG
1201 AGGAACGGTC CAGTGCAGAT CAGCTCAACA GCCAAGAAGA GCTTACTCTT
1251 TCCAACCCAG ACCTCTTCGA CTGCCCCAAGC CCACAGATTG CCAAAGTGGT
1301 ATTGCTTTCC AAAGTGCCTC AGTGGGGCAG CCTCTGAGAA CTGTTCAAAA
1351 GGACAGTAGT AGCAGATCAA AGAAAACTT CTATAGAAAT GAAACCTCCA
1401 GCCAGGAGTC TAAGAACTTG TCCACACCAG GAACCAAGT TCAGGCCCGA
1451 GGAAGAATCC TACCTGGTTC CCCTGTGAAG AGAACCTGGC ACCGACATCT
1501 TAAAGACAAA CTCACACACA AGGAGCATAA CCACCCAGC TTCTATAGGG
1551 AGAGAACCCC ACGCGGTCCT TCTGAGAGAA CCCGTCATAA CCCCTCTTGG
1601 AGAAACCATC GCAGTCCCTC TGAGAGAAGC CAACGCAGTT CCTTGGAGAG
1651 AAGACATCAC AGTCCCTCTC AGAGGAGCCA CTGCAGTCCC TCTAGGAAAA
1701 ACCATTCCAG TCCTTCTGAG AGAAGCTGGC GCAGTCCGTC TCAGAGAAAT
1751 CACTGCAGTC CCCCCGAGAG GAGCTGTAC AGTCTCTCTG AAAGGGGCCT
1801 TCACAGTCCC TCTCAGAGGA GCCATCGCGG TCCCTCTCAG AGAAGACATC
1851 ACAGTCCCTC AGAGAGAAGC CATCGCAGTC CCTCAGAGAG AAGCCATCGC
1901 AGTCCCTCTG AGAGAAGACA TCGCAGTCCC TCCCAGAGGA GCCATCGCGG
1951 TCCCTCAGAG AGAAGCCATT GCAGTCCCTC TGAGAGAAGA CATCGCAGTC
2001 CCTCTCAGAG GAGCCATCGT GGTCCCTCTG AGAGAAGACA TCCAGTCCC
2051 TCTAAGAGAA GCCATCGCAG TCCCGCTCGG AGGAGCCATC GCAGTCCCTC
2101 AGAGAGAAGC CATCACAGTC CCTCTGAGAG AAGCCATCAC AGTCCCTCTG
2151 AGAGAAGACA TCACAGTCCC TCTGAGAGAA GCCATTGCAG TCCCTCTGAG
2201 AGAAGCCATT GCAGTCCCTC TGAGAGAAGA CATCGCAGTC CCTCTGAGAG
2251 AAGACATCAC AGTCCCTCAG AGAAAAGCCA TCACAGTCCC TCTGAGAGAA
2301 GCCATCACAG TCCCTCTGAG AGAAGACGTC ACAGTCCCTT GGAGAGGAGC
2351 CGTCACAGTC TCTTGGAGAG GAGCCATCGC AGTCCCTCTG AGAGGAGATC
2401 TCACAGGTCC TTTGAGAGGA GCCATCGTAG GATTCTGAG AGAAGTCACA
2451 GTCCCTCAGA GAAGAGCCAC CTCAGTCCCT TGGAAAGAAG CCGTTGCAGT
2501 CCCTCTGAGA GGAGAGGACA CAGTTCCCTT GGGAAAACCT GTCACAGTCC
2551 CTCTGAGAGA AGCCATCGCA GTCCCTCCGG GATGAGGCAA GGGAGGACCT
2601 CTGAGAGGAG CCATCGCAGT TCCTGTGAGA GAACCCGTCA CAGTCCCTCT
```

```

2651 GAGATGAGGC CAGGGAGGCC CTCTGGGAGG AACCATTGCA GTCCCTCTGA
2701 GAGGAGCCGA CGCAGTCCCC TTAAGGAGGG ACTCAAGTAC AGTTTCCCTG
2751 GAGAGAGGCC CAGCCATAGT TTGTCTAGAG ATTTCAGAA TCAAACAACT
2801 CTCCTCGGGA CCACACATAA AAATCCCAA GCAGGGCAAG TGTGGAGGCC
2851 TGAAGCTACT CGATGAGGCG AGGTCCGCCC CTATTATTCA TTGTCTAAG
2901 TCTTCATCGT GCTGCCCTTT CCAGGCTTCT TTCCTGCTCA GCCACTGCCT
2951 CCAATTCTGT CGCCCCAGC GTGGAAAGGC TTCCATTTCT CTCTACCGGG
3001 GGGGAGGCGG GTGAGAATGG GTCTGTAATT TCTCTAAGAT GAATAAAGGG
3051 GCAGTTAATT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 2863 bp; peptide length: 939
 Category: similarity to unknown protein
 Classification: unclassified
 Prosite motifs: ATP_GTP_A (824-832)

```

1 MEESEDSQSD SQTRISESQH SLKPNYLSQA KTDSEQFQL LEDLQLKIAA
51 KLLRSQIPPD VPPPLASGLV LKYPICLQCG RCSCGLNCHHK LQTTSGPYLL
101 IYPQLHLVLT PEGHGEVRLH LGFRLRIGKR SQISKYRERD RPYIRRSPI
151 PSQRKAKIYT QASKSPTSTI DLQSGPSQSP APVQVYIRRG QRSRFDLVEK
201 TKTRAPGHYE FTQVHNLPEP DSESTQNEKR AKVRTKKTSD SKYPMKRITK
251 RLKHKRKFYT NSRTTIESPS RELAAHLRRK RIGATQTSTA SLKRQPKKPS
301 QPKFMQLLFQ SLKRAFQTAH RVIASVGRKP VDGTRPDNLW ASKNYYPKQN
351 ARDYCLPSSI KRDKRSADKL TPAGSTIKQE DILWGGTVQC RSAQQPRRAY
401 SFQPRPLRLP KPTDSQSGIA FQTASVGQPL RTVQKDSSSR SKKNFYRNET
451 SSQESKNLST PGTRVQARGR ILPGSPVKRT WRRHLKDKLT HKEHNHPSFY
501 RERTPRGPSE RTRHNPSWRN HRSPERSQR SSLERRHHSP QRSRSHCSPSR
551 KNHSSPSERS WRSPSQRNHC SPPERSCHSL SERGLHSPSQ RSHRGPSQRR
601 HHSPERSHR SPERSHRSP SERRHRSPSQ RSHRGPSERS HCSPSERRHR
651 SPSQRSHRGP SERRHHSPPK RSHRSPARRS HRSPERSHH SPERSHHSP
701 SERRHHSPE RSHCSPERS HCSPSERRHR SPERRHHSP SEKSHHSPSE
751 RSHHSPSERR RSHPLERSRH SLLERSHRSP SERRSHRSFE RSHRRISERS
801 HSPSEKSHLS PLERSRCSPE ERGRHSSSGK TCHSPERSH RSPSGMRQGR
851 TSERSHRSSC ERTRHSPSEM RPRGPRGRNH CSPSERSRRS PLKEGLKYSF
901 PGERPSHSLR RDFKNQTLL GTTHKNPKAG QVWRPEATR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8g11, frame 2

TREMBL:AF061185_1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; Phytophthora infestans cyst germination specific acidic repeat protein precursor (car90) gene, complete cds., N = 1, Score = 457, P = 2.3e-39

TREMBL:AC004561_38 gene: "F16P2.41"; product: "putative proline-rich protein"; Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence, complete sequence., N = 1, Score = 340, P = 4.2e-27

TREMBL:AF062655_1 product: "plenty-of-prolines-101"; Mus musculus plenty-of-prolines-101 mRNA, complete cds., N = 1, Score = 313, P = 3.6e-24

PIR:PN0099 son3 protein - human (fragment), N = 1, Score = 292, P = 1.2e-22

>TREMBL:AF061185_1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; Phytophthora infestans cyst germination specific acidic repeat protein precursor (car90) gene, complete cds.

Length = 1,489

HSPs:

Score = 457 (68.6 bits), Expect = 2.3e-39, P = 2.3e-39
 Identities = 91/444 (20%), Positives = 239/444 (53%)

Query: 475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPGSERTRHNPSWRNHRSPSERSQRSSL 533
 +P + T + +++ T+ ++ E TP P+E T + P+ +P+E + +S
 Sbjct: 584 APTEETMYAPIET-TYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAST 642

Query: 534 ERRHHSQSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQSRSH 593
 E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +
 Sbjct: 643 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 702

Query: 594 RGPQRHHSPSERSHRSPSERSHRSPSERRHRSPSQSRHSGPSERSHCSPSERRHRSPS 653
 P++ + P+E + +P+E + +P+E +P + + GP+E + +P+E +P+
 Sbjct: 703 YAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIETTYGPTTEETTYAPTEATTYAPT 762

Query: 654 QRSHRGPSERRHHSKRSRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSH 713
 + + P+E + P+ + +P + +P+E + ++P+E + ++P+E + P+E +
 Sbjct: 763 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETT 822

Query: 714 CSPSERSHCSPSERRHRSPSERRHHSPEKSHHSPSERSHHSPSERRHHSPLERSRHSLL 773
 +P+E + P+E +P+E ++P+E++ ++P+E++ ++P+E ++P E + +
 Sbjct: 823 YAPTEETPYEPTTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPT 882

Query: 774 ERSRHSRSPERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTC 832
 E + +P++ ++ E + + E ++P+E++ +P E + P+E ++ + +T
 Sbjct: 883 EETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 942

Query: 833 HSPSERSHRSPSGMRQGRTERSRRSCERTRHSPEMRPGRPSGRNHCSPSERSRRSPL 892
 ++P+E + +P+ +E + + E T + P+E P+ +P+E + +P+
 Sbjct: 943 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPI 1002

Query: 893 KEGLKYSFPGERPSSLSRDFKNQTT 918
 +E Y+ P E +++ + + + T
 Sbjct: 1003 EE-TTYA-PTEETTYAPAEETPYEPT 1026

Score = 445 (66.8 bits), Expect = 4.5e-38, P = 4.5e-38
 Identities = 83/394 (21%), Positives = 212/394 (53%)

Query: 502 ERTPRGPGSERTRHNPSWRNHRSPSERSQRSSLERRHHSQSRSHCSPSRKNHSSPSERSW 561
 E TP P+E T + P+ +P+E + + E ++P++ + +P+ + P+E +
 Sbjct: 763 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETT 822

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSRHSGPSQRHHSPSERSHRSPSERSHRSPS 621
 +P++ P E + ++ +E ++P++ + P+++ ++P+E + +P+E + P+
 Sbjct: 823 YAPTEETPYEPTTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPT 882

Query: 622 ERRHRSQSRSHRGPSERSHCSPSERRHRSPSQSRHSGPSERRHHSKRSRSPARRSH 681
 E +P++ + P+E + + +E +P++ + P+E + P++ + +P +
 Sbjct: 883 EETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 942

Query: 682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHS 741
 +P+E + ++P+E + ++P+E ++P+E + P+E + +P+E +P+E ++P
 Sbjct: 943 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPI 1002

Query: 742 EKSHHSPSERSHHSPSERRHHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ ++P+E + ++P+E + P E + ++ E + +P+E ++ S E + + E +
 Sbjct: 1003 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETT 1062

Query: 801 HSPSEKSHLSPLERSRCSPEERRGHSSSGKTC HSPSERSHRSPSGMRQGRTERSRRSSC 860
 ++P+E++ P E + +P+E ++ + +T ++P+E + +P+ +E +
 Sbjct: 1063 YAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT 1122

Query: 861 ERTHRSPSEMRPGRPSGRNHCSPSERSRRSPLKE 894
 E T ++P+E P+ +P E + P +E
 Sbjct: 1123 EETTYAPTEETTYAPTEETMYAPIETTYGPTTE 1156

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
 Identities = 86/421 (20%), Positives = 223/421 (52%)

Query: 475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPGSERTRHNPSWRNHRSPSERSQRSSL 533
 +P + T + +K T+ ++ E TP P+E T + P+ +P+E + +S
 Sbjct: 848 APTEETTYAPT-EKTTYAPTEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYAST 906

Query: 534 ERRHHSQSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQSRSH 593
 E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +
 Sbjct: 907 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 966

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
Identities = 91/434 (20%), Positives = 232/434 (53%)

Score = 437 (65.6 bits), Expect = 3.3e-37, P = 3.3e-37
Identities = 85/417 (20%), Positives = 223/417 (53%)

969

Query: 801 HSPSEKSHLSPLERSRCSPEERRGHSSSGKTCCHSPSERSHRSPSGMRQGRTERSRRSSC 860
 ++P+E++ +P E + +P E + + +T ++P+E + +P+ +E +
 Sbjct: 719 YAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPT 778

Query: 861 ETRHSPSEMRPGRPSGRNHCSPEERSRRSPLKEGLKYSFPGERPSHLSRDFKNQTT 918
 T ++P+E P+ +P+E + +P +E Y P E +++ + + T
 Sbjct: 779 GETTYAPTEETTYAPTEETTYAPTEETTYAPTEE-TPYE-PTEETTYAPTEETPYEPT 834

Score = 428 (64.2 bits), Expect = 3.1e-36, P = 3.1e-36
 Identities = 89/440 (20%), Positives = 228/440 (51%)

Query: 473 PGSPVKRTWHRHLKDKLTHKEHNHPSFYR-ERTPRGPSETRHNPWRNHRSPEERSQRS 531
 P P + T + K+ T+ ++ E T P+E T + P+ P+E + +
 Sbjct: 470 PYEPTTEETTYAPTKET-TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYA 528

Query: 532 SLERRHSPSQSHCSPESRKNHSSPERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQS 591
 E ++P++ + +P+ + +P+E + +P++ P E + ++ +E ++P++
 Sbjct: 529 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEE 588

Query: 592 SHRGPSQRRHSPSERSHRSPSERSHRSPSERRHRSPEQRSHRGPSERSHCSPEERRHRS 651
 + P + ++P+E + +P+E + P+E +P++ + P+E + + +E +
 Sbjct: 589 TMYAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYA 648

Query: 652 PSQSRHSGPSERRHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHSPSER 711
 P++ + P+E + P++ + +P + +P+E + ++P+E + ++P+E ++P+E
 Sbjct: 649 PTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 708

Query: 712 SHCSPEERSHCSPEERRHSPSERRHSPSEKSHHSPSERSHHSPSERRHSPSERSRHS 771
 + P+E + +P+E +P+E ++P E++ + P+E + ++P+E ++P E + ++
 Sbjct: 709 TPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYA 768

Query: 772 LLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPEERRGHSSSGK 830
 E + P+ ++ E + + E +++P+E++ +P E + P+E ++ + +
 Sbjct: 769 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEE 828

Query: 831 TCHSPSERSHRSPSGMRQGRTERSRRSSCERTRHSPSEMRPGRPSGRNHCSPEERSRRS 890
 T + P+E + +P+ +E + + E+T ++P+E P+ P+E + +
 Sbjct: 829 TPYEPTTEETTYPTTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTTEETTYA 888

Query: 891 PLKEGLKYSFPGERPSHLSRD 912
 P KE Y+ P E +++ + +
 Sbjct: 889 PTKE-TTYA-PTEETTYASTE 908

Score = 427 (64.1 bits), Expect = 4.0e-36, P = 4.0e-36
 Identities = 81/394 (20%), Positives = 213/394 (54%)

Query: 502 ERTPRGPSETRHNPWRNHRSPEERSQSSLERRHSPSQSHCSPESRKNHSSPERSW 561
 E T GP+E T + P+ +P+E + + E + P+ + +P+ + +P+E +
 Sbjct: 739 EETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 798

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHRSPSERSHRSPS 621
 +P++ +P E + + +E ++P++ + P++ ++P+E + +P+E + +P+
 Sbjct: 799 YAPTEETTYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETTYAPT 858

Query: 622 ERRHRSPEQRSHRGPSERSHCSPEERRHRSPEQRSHRGPSERRHSPSKRSHRSPARRSH 681
 E+ +P++ + P+E + P+E +P++ + P+E ++ ++ +P +
 Sbjct: 859 EKTYYAPTEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYASTEETTYAPTEETT 918

Query: 682 RSPSERSHHSPSERSHHSPSERRHSPSERSHCSPEERSHCSPEERRHRSPEERRHSPS 741
 +P+E + + P+E + ++P+E ++P+E + +P+E + +P+E +P+E + P+
 Sbjct: 919 YAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT 978

Query: 742 EKSHHSPSERSHHSPSERRHSPSERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ ++P+E + ++P+E ++P+E + ++ E + +P+E + E + + E +
 Sbjct: 979 EETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 1038

Query: 801 HSPSEKSHLSPLERSRCSPEERRGHSSSGKTCCHSPSERSHRSPSGMRQGRTERSRRSSC 860
 ++P+E++ + E + +P+E ++ + +T + P+E + +P+ +E + +
 Sbjct: 1039 YAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPT 1098

Query: 861 ETRHSPSEMRPGRPSGRNHCSPEERSRRSPLKE 894
 E T ++P+E P+ P+E + +P +E
 Sbjct: 1099 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEE 1132

Score = 424 (63.6 bits), Expect = 8.5e-36, P = 8.5e-36
 Identities = 81/394 (20%), Positives = 210/394 (53%)

Query: 502 ERTPRGPSETRHNPWRNHRSPEERSQSSLERRHSPSQSHCSPESRKNHSSPERSW 561
 E T P+E T + P+ +P+E + + E + P++ + +P+ + +P+E +
 Sbjct: 939 EETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETM 998

Query: 562 RSPSQRNHCSPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHRSPSERSHRSPS 621
 +P + +P E + ++ +E + P++ + P++ ++P+E + + +E + +P+
 Sbjct: 999 YPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPT 1058

Query: 622 ERRHRSPSQSRHSGPSERSHCSPSERRHRSPSQSRHSGPSERRHSPSKRSHRSPARRSH 681
 E +P++ + P+E + +P+E +P++ + P+E ++P++ + +PA +
 Sbjct: 1059 EETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 1118

Query: 682 RSPSERSHHSPSERSHHSPSERRHSPSERSHCSPSERSHCSPSERRHRSPSERRHSPS 741
 P+E + ++P+E + ++P+E ++P E + P+E + +P+E +P+E ++P+
 Sbjct: 1119 YEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEETTYAPTEETPYAPT 1178

Query: 742 EKSHHSPSERSHHSPSERRHSPSLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ + P+ + ++P+E ++P E + ++ E + +P+E + E + + E +
 Sbjct: 1179 EETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETT 1238

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTERSRRSSC 860
 + P+E++ +P E + +P+E ++ + +T ++P + + P+ +E + +
 Sbjct: 1239 YEPTTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTTEETTYAPTEETTYAPT 1298

Query: 861 ETRHSPSEMRPGRPSGRNHCSPSERSRRSPLKE 894
 E T ++P+E P+G +P+E + +P +E
 Sbjct: 1299 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEE 1332

Score = 422 (63.3 bits), Expect = 1.4e-35, P = 1.4e-35
 Identities = 84/407 (20%), Positives = 216/407 (53%)

Query: 502 ERTPRGSPERTRHNPWRNHRSERSQRSSLERRHSPSQSRHCSPSRKNHSSPSERSW 561
 E T P+E T + P+ P+E + + E + P++ + +P+ + +P+E +
 Sbjct: 795 EETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETT 854

Query: 562 RSPSQRNHCSPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHRSPSERSHRSPS 621
 +P+++ +P E + ++ +E + P++ + P++ ++P+E + + +E + +P+
 Sbjct: 855 YAPTEKTTYAPTEETTYAPTEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPT 914

Query: 622 ERRHRSPSQSRHSGPSERSHCSPSERRHRSPSQSRHSGPSERRHSPSKRSHRSPARRSH 681
 E +P++ + P+E + +P+E +P++ + P+E ++P++ + +PA +
 Sbjct: 915 EETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 974

Query: 682 RSPSERSHHSPSERSHHSPSERRHSPSERSHCSPSERSHCSPSERRHRSPSERRHSPS 741
 P+E + ++P+E + ++P+E ++P E + +P+E + +P+E P+E ++P+
 Sbjct: 975 YEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPT 1034

Query: 742 EKSHHSPSERSHHSPSERRHSPSLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ ++P+E + ++ +E ++P E + ++ E + P+E ++ E + + E +
 Sbjct: 1035 EETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETT 1094

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTERSRRSSC 860
 ++P+E++ +P E + +P+E + + +T ++P+E + +P+ E +
 Sbjct: 1095 YAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPT 1154

Query: 861 ETRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHS 908
 E T ++P+E P+ +P+E + P E Y+ P E +++
 Sbjct: 1155 EETTYAPTEATTYAPTEETPYAPTEETTYEPTGE-TTYA-PTEETTYA 1200

Score = 421 (63.2 bits), Expect = 1.8e-35, P = 1.8e-35
 Identities = 86/418 (20%), Positives = 219/418 (52%)

Query: 491 HKEHNHSPFYRETRPRGSPERTRHNPWRNHRSERSQRSSLERRHSPSQSRHCSPSR 550
 H H E T P+E T + P+ +P+E + + E + P++ + +P+
 Sbjct: 376 HYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYTPT 435

Query: 551 KNHSSPSERSWRSPSQRNHCSPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHR 610
 + +P+E + +P+++ +P E + ++ +E + P++ + P++ ++P+E +
 Sbjct: 436 ETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTTEETTYAPTETTYAPTEETTY 495

Query: 611 SPERSHRSPSERRHRSPSQSRHSGPSERSHCSPSERRHRSPSQSRHSGPSERRHSPSK 670
 + +E + +P+E +P++ + P+E + +P+E +P++ + P+E ++P++
 Sbjct: 496 ASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 555

Query: 671 RSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHSPSERSHCSPSERSHCSPSERRHR 730
 + +PA + P+E + ++P+E + ++P+E ++P E + +P+E + +P+E
 Sbjct: 556 ETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPY 615

Query: 731 SPERRHSPSEKSHHSPSERSHHSPSERRHSPSLERSRHSLLERSHRSPSERRSHRSFE 790
 P+E ++P+E++ ++P+E + ++ +E ++P E + ++ E + P+E ++ E
 Sbjct: 616 EPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTE 675

Query: 791 RS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQ 849
 + + E +++P+E++ +P E + +P+E + + +T ++P+E + +P+

Sbjct: 676 ETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMY 735

Query: 850 RTSERSHRSSCERTRHNSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHS 908
E + E T ++P+E P+ +P+E + P E Y+ P E +++
Sbjct: 736 APIEETTYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPTGE-TTYA-PTEETTYA 792

Score = 420 (63.0 bits), Expect = 2.3e-35, P = 2.3e-35
Identities = 82/393 (20%), Positives = 206/393 (52%)

Query: 502 ERTPRGSPERTRHNPSPWRNHRSPPERSQSSSLERRHHSPSQSHCSPSRKNHSSPERSW 561
E TP P+E T + P+ +P+E + + +E ++P++ + +P+ + P+E +

Sbjct: 971 EETPYEPTETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTET 1030

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSHRGPSQRRHHSPERSHRSPERSHRSPS 621
+P++ +P E + ++ +E ++P++ + P++ + P+E + +P+E + +P+

Sbjct: 1031 YAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPT 1090

Query: 622 ERRHRSPSQSHRGPSERSHCSPSERRHRSPSQSHRGPSERRHHSPSKRSHRSPARRSH 681
E +P++ + P+E + +P+E P++ + P+E ++P++ + +P +

Sbjct: 1091 EETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIEETT 1150

Query: 682 RSPERSHHSPERSHHSPERRHHSPERSHCSPSERSHCSPSERRHRSPERRHHSPS 741
P+E + ++P+E + ++P+E ++P+E + P+ + +P+E +P+E ++P+

Sbjct: 1151 YGPTETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPT 1210

Query: 742 EKSHHSPERSHHSPERRHSPERSHLLERSHRSPERRSHRSFERS-HRRISERS 800
E++ ++P+E + + P+E ++P E + + E + +P+E ++ E + + E

Sbjct: 1211 EETTYAPTEETPYEPTETTYAPTEETTYEPTETTYAPTEETTYAPTEETTYAPTEETM 1270

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCCHSPERSHRSPSGMRQRTSERSHRSSC 860
++P +++ P E + +P+E ++ + +T ++P+E + P+G +E + +

Sbjct: 1271 YAPIDETTYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPT 1330

Query: 861 ETRHSPSEMRPGRP-----SGRNHCSPE 885
E T ++P E P P S C+ E

Sbjct: 1331 EETTYAPMEETPYEPAEESTSTVSTKPCNTEE 1363

Score = 419 (62.9 bits), Expect = 3.0e-35, P = 3.0e-35
Identities = 83/411 (20%), Positives = 215/411 (52%)

Query: 502 ERTPRGSPERTRHNPSPWRNHRSPPERSQSSSLERRHHSPSQSHCSPSRKNHSSPERSW 561
E T P+E T + P+ +P+E + E ++P++ + +P+ + +P E +

Sbjct: 947 EETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIEETT 1006

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSHRGPSQRRHHSPERSHRSPERSHRSPS 621
+P++ +P E + + +E ++P++ + P++ ++ +E + +P+E + +P+

Sbjct: 1007 YAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 1066

Query: 622 ERRHRSPSQSHRGPSERSHCSPSERRHRSPSQSHRGPSERRHHSPSKRSHRSPARRSH 681
E P++ + P+E + +P+E +P++ + P+E ++P++ + P +

Sbjct: 1067 EETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETT 1126

Query: 682 RSPERSHHSPERSHHSPERRHHSPERSHCSPSERSHCSPSERRHRSPERRHHSPS 741
+P+E + ++P+E + ++P E + P+E + +P+E + +P+E +P+E + P+

Sbjct: 1127 YAPTEETTYAPTEETMYAPIEETTYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPT 1186

Query: 742 EKSHHSPERSHHSPERRHSPERSHLLERSHRSPERRSHRSFERS-HRRISERS 800
++ ++P+E + ++P+E ++P E + ++ E + P+E ++ E + + E +

Sbjct: 1187 GETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAPTEETTYEPTETT 1246

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCCHSPERSHRSPSGMRQRTSERSHRSSC 860
++P+E++ +P E + +P+E ++ +T + P+E + +P+ +E + +

Sbjct: 1247 YAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTETTYAPTEATTYAPTEETPYAPT 1306

Query: 861 ETRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHSLSRD 912
E T + P+ P+ +P+E + +P++E Y P E + ++S +

Sbjct: 1307 EETTYEPTGETTYAPTEETTYAPTEETTYAPMEE-TPYE-PAEESTSTVSTE 1356

Score = 415 (62.3 bits), Expect = 8.0e-35, P = 8.0e-35
Identities = 84/423 (19%), Positives = 218/423 (51%)

Query: 473 PGSPVKRTWHRHLKDKLTHKEHNHPSFYR-ERTPRGSPERTRHNPSPWRNHRSPPERSQRS 531
P P + T + K+ T+ ++ E T P+E T + P+ P+E + +

Sbjct: 878 PYEPTETTYAPTKET-TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTETTYA 936

Query: 532 SLERRHHSPSQSHCSPSRKNHSSPERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQ 591
E ++P++ + +P+ + +P+E + +P++ P E + ++ +E ++P++

Sbjct: 937 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEE 996

Query: 592 SHRGPSQRRHHSPERSHRSPERSHRSPERRHRSPSQSHRGPSERSHCSPSERRHRS 651
+ P + ++P+E + +P+E + P+E +P++ + P+E + + +E +

Sbjct: 997 TMYAPIETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYASTEETTYA 1056

Query: 652 PSQRSHRGPSERRHHSPSKRSHRSPARRSHRSPERSHHSPERSHHSPERRHHSER 711
P++ + P+E + P++ + P+ + P+E + ++P+E + ++P+E ++P+E

Sbjct: 1057 PTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 1116

Query: 712 SHCSPERSHCSPEERRHRSPEERRHSPSEKSHHSPERSHHSPERRHSPERSHSPERSH 771
+ P+E + +P+E +P+E ++P E++ + P+E + ++P+E ++P E + ++

Sbjct: 1117 TPYEPTETTYAPTEETTYAPTEETMYAPIEETTYGPTETTYAPTEATTYAPTEETPYA 1176

Query: 772 LLERSHRSPERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPEERRGHSSSGK 830
E + P+ ++ E + + E +++P+E++ +P E + P+E ++ + +

Sbjct: 1177 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAPTEE 1236

Query: 831 TCHSPERSHRSPSGMRQRTSERSHRSSCERTRHSPSEMGRPSGRNHCSPEERSRRS 890
T + P+E + +P+ +E + + E T ++P + P+ +P+E + +

Sbjct: 1237 TTYEPTETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTETTYAPTEATTYA 1296

Query: 891 PLKE 894
P +E

Sbjct: 1297 PTEE 1300

Score = 403 (60.5 bits), Expect = 1.6e-33, P = 1.6e-33
Identities = 84/394 (21%), Positives = 213/394 (54%)

Query: 501 RERTPRGPSETRHNPSWRNHRSPEERSQSSSLERRHHSQQRSHCSPEKSHHSPERS 560
RE T PSE T + P +P+E+ +E + + ++ +P++ ++P+ER

Sbjct: 319 REETTAAPSEDITYAPREVTPYAPTEKPY--DVEETTYVTEESTY-APTKSETNAPTERM 375

Query: 561 WRSPQRNHCSPPERSCHSLSERGLHSPSQSRHSGPSQRRHHSPEERSHRSPERSHRSP 620
+ ++ C E + ++ +E ++P++ + P++ ++P+E + P+E + +P

Sbjct: 376 HYAHIEKP-CDT-EVTMYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYTP 433

Query: 621 SERRHRSPQRSHRGPSERSHCSPEERRHRSPEERSHSGPSERSHCSPEERRHRSPEERRHSP 680
+E +P++ + P+E++ +P+E +P++ + P+E ++P+K + +P +

Sbjct: 434 TEETTYAPTEETTYAPTEKTYAPTEETTYAPTEETPYEPTETTYAPTKETTYAPTEET 493

Query: 681 HRSPERSHHSPERSHHSPERRHHSPEERSHCSPEERSHCSPEERRHRSPEERRHSP 740
+ +E + ++P+E + ++P+E + P+E + +P+E + +P+E +P+E ++P

Sbjct: 494 TYASTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAP 553

Query: 741 SEKSHHSPERSHHSPERRHSPERSHLSPLERSHRSPEERSHRSFERS-HRRISER 799
+E++ ++P+E + + P+E ++P E + ++ E + +P E ++ E + + E

Sbjct: 554 TEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEET 613

Query: 800 SHSPSEKSHLSPLERSRCSPEERRGHSSSGKTCHSPERSHRSPSGMRQRTSERSHRSS 859
+ P+E++ +P E + +P+E ++S+ +T ++P+E + +P+ +E + +

Sbjct: 614 PYEPTETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTETTYAP 673

Query: 860 CERTHSPSEMGRPSGRNHCSPEERSRRSPLKE 894
E T ++P+E P+ +P+E + +P +E

Sbjct: 674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 708

Score = 398 (59.7 bits), Expect = 5.5e-33, P = 5.5e-33
Identities = 84/402 (20%), Positives = 209/402 (51%)

Query: 475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSETRHNPSWRNHRSPEERSQSSSL 533
+P + T + +++ T+ ++ E TP P+E T + P+ +P+E + +S

Sbjct: 992 APTEETMYAPIEET-TYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAST 1050

Query: 534 ERRHHSQQRSHCSPEKSHHSPERSHRSPEERSHRSPEERSHRSPEERSHRSPEERSH 593
E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +

Sbjct: 1051 EETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 1110

Query: 594 RGPSQRRHHSPEERSHRSPERSHRSPEERRHRSPEERSHSGPSERSHCSPEERRHRS 653
P++ + P+E + +P+E + +P+E +P + + GP+E + +P+E +P+

Sbjct: 1111 YAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIEETTYGPTETTYAPTEATTYAPT 1170

Query: 654 QRSHRGPSERRHHSPEKSHRSPARRSHRSPERSHHSPERSHHSPERRHHSPEERSH 713
+ + P+E + P+ + +P + +P+E + ++P+E + ++P+E +P+E +

Sbjct: 1171 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTET 1230

Query: 714 CSPERSHCSPEERRHRSPEERRHSPSEKSHHSPERSHHSPERRHSPERSHLSPLERSHLL 773
+P+E + P+E +P+E ++P+E++ ++P+E + ++P + + P E + ++

Sbjct: 1231 YAPTEETTYEPTETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTETTYAPT 1290

Query: 774 ERSHRSPERRSHRSFERSHRRISERSHSPSEKSHLSPLERSRCSPEERRGHSSSGKTCH 833
E + +P+E + E E ++ P+ ++ +P E + +P+E ++ +T +

Sbjct: 1291 EATTYAPTEETPYAPTE-----ETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPY 1343

Query: 834 SPERSHRSPSGMRQRTSERSHRSSCERTRHSPSEMGRPS 876

P+E S + S + T E + + E T PS+ P+
 Sbjct: 1344 EPAEESTSTVSTKPCNTEFTDEPTDEPT-DEPSDEPTDEPT 1385

Score = 368 (55.2 bits), Expect = 9.5e-30, P = 9.5e-30
 Identities = 79/386 (20%), Positives = 211/386 (54%)

Query: 524 PSERSQSSLERRHSPSQSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSER 583
 PS+ ++ + E + P + + +PS +P E + +P+++ + E + + ++E
 Sbjct: 303 PSDETEAPT-EGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPY--DVEETTY-VTEE 358

Query: 584 GLHSPSQSRSHRGPSQRRHSPSER-----SHRSPSERSHRSPSERRHSPSQSRSHRGPS 637
 ++P++ P++R H++ E+ + +P+E + +P+E +P++ + P+
 Sbjct: 359 STYAPTKESTNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPT 418

Query: 638 ERSCHSPSERRHSPSQSRSHRGPSERRHSPSKRSHRSPARRSHRSPSERSHHSPSERSH 697
 E + P+E +P++ + P+E ++P+++ +P + +P+E + + P+E +
 Sbjct: 419 EETPYEPTETTYPTTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTETT 478

Query: 698 HSPSERRHSPSERSHCSPSERSHCSPSERRHSPSERRHSPSEKSHHSPSERSHHSPS 757
 ++P++ ++P+E + + +E + +P+E +P+E + P+E++ ++P+E + ++P+
 Sbjct: 479 YAPTKEETTYAPTEETTYASTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPT 538

Query: 758 ERRHSPSLERSRHSLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSR 816
 E ++P E + ++ E + +P+E + E + + E ++P+E++ +P+E +
 Sbjct: 539 EETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIEETT 598

Query: 817 CSPSERRGHSSSGKTCHSPSERSHRSPSGMRQRTSERSHRSSCERTRHSPSEMRPGRPS 876
 +P+E ++ + +T + P+E + +P+ +E + +S E T ++P+E P+
 Sbjct: 599 YAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 658

Query: 877 GRNHCSPSERSRRSPLKEGLKYSFPGERPSHS 908
 P+E + +P +E Y+ P E +++
 Sbjct: 659 EETPYEPTETTYAPTEE-TTYA-PTEETTYA 688

Score = 337 (50.6 bits), Expect = 2.1e-26, P = 2.1e-26
 Identities = 66/328 (20%), Positives = 170/328 (51%)

Query: 502 ERTPRGPSETRHNPSWRNHRSPPERSQSSLERRHSPSQSRSHCSPSRKNHSSPSERSW 561
 E T P+E T + P+ +P+E + + E ++P++ + +P+ + +P+E +
 Sbjct: 1059 EETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 1118

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSRSHRGPSQRRHSPSERSHRSPSERSHRSPS 621
 P++ +P E + ++ +E +++P + + GP++ ++P+E + +P+E + +P+
 Sbjct: 1119 YEPTETTYAPTEETTYAPTEETMYAPIEETTYGPTETTYAPTEATTYAPTEETPYAPT 1178

Query: 622 ERRHRSQSRSHRGPSERSHCSPSERRHRSQSRSHRGPSERRHSPSKRSHRSPARRSH 681
 E P+ + P+E + +P+E +P++ + P+E + P++ + +P +
 Sbjct: 1179 EETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAPTEETT 1238

Query: 682 RSPSERSHHSPSERSHHSPSERRHSPSERSHCSPSERSHCSPSERRHRSPPERSHHSPS 741
 P+E + ++P+E + ++P+E ++P+E + +P + + P+E +P+E ++P+
 Sbjct: 1239 YEPTETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTETTYAPTEATTYAPT 1298

Query: 742 EKSHHSPSERSHHSPSERRHSPSLERSRHSLERSHRSPSERRSHRSFERSHRRIS---- 797
 E++ ++P+E + + P+ ++P E + ++ E + +P E + E S +S
 Sbjct: 1299 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPYEPAEESTSTVSTKEP 1358

Query: 798 ----ERSHSPSEKSHLSPLERSRCSPE 821
 E + P+++ P + P++
 Sbjct: 1359 CNTEFTDEPTDEPTDEPSDEPTDEPTD 1386

Score = 333 (50.0 bits), Expect = 5.7e-26, P = 5.7e-26
 Identities = 63/320 (19%), Positives = 166/320 (51%)

Query: 502 ERTPRGPSETRHNPSWRNHRSPPERSQSSLERRHSPSQSRSHCSPSRKNHSSPSERSW 561
 E T P+E T + P+ +P+E + + E ++P++ + P+ + +P+E +
 Sbjct: 1075 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETT 1134

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSRSHRGPSQRRHSPSERSHRSPSERSHRSPS 621
 +P++ +P E + + +E ++P++ + P++ ++P+E + P+ + +P+
 Sbjct: 1135 YAPTEETMYAPIEETTYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPT 1194

Query: 622 ERRHRSQSRSHRGPSERSHCSPSERRHRSQSRSHRGPSERRHSPSKRSHRSPARRSH 681
 E +P++ + P+E + +P+E P++ + P+E + P++ + +P +
 Sbjct: 1195 EETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAPTEETTYEPTETTYAPTEETT 1254

Query: 682 RSPSERSHHSPSERSHHSPSERRHSPSERSHCSPSERSHCSPSERRHRSPPERSHHSPS 741
 +P+E + ++P+E + ++P + + P+E + +P+E + +P+E +P+E + P+
 Sbjct: 1255 YAPTEETTYAPTEETMYAPIDETYYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPT 1314

Query: 742 EKSHHSPSERSHHSPSERRHSPSLERSRHSLERSHRSPSERRSHRSFERSHRRISERSH 801

++ ++P+E + ++P+E ++P+E + + E S + S + + E + E +
 Sbjct: 1315 GETTYAPTEETTYAPTEETTYAPMEETPYEPAEESTSTVSTKPCNTEETDEPTDEPTD 1374
 Query: 802 SPSEKSHLSPLERSRCSPE 821
 PS++ P + P++
 Sbjct: 1375 EPSDEPTDEPTDEPTDLPTD 1394

Score = 303 (45.5 bits), Expect = 9.6e-23, P = 9.6e-23
 Identities = 70/322 (21%), Positives = 170/322 (52%)

Query: 584 GLHSPSQSRHSGPSQRRHSPSERSHRSPSERSHRSPSERRHRSQSRHSGPSERSHCS 643
 G + PS + P++ + P E + +PSE + +P E +P+++ + E ++ +
 Sbjct: 299 GGYEPSDETE-APTEGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPY-DVEETTYVT 356
 Query: 644 PSERRHRSQSRHSGPSERRHSPSKRSHRSPARRSHRSPERSHHSPERSHHSPSER 703
 E +P++ P+ER H++ ++ + + +P+E + ++P+E + ++P+E
 Sbjct: 357 --EESTYAPTKESETNAPTERMHYAHIEKPCDTEV--TMYAPTEETTYAPTEETTYAPTEE 412
 Query: 704 RHSPSERSHCSPSERSHCSPSERRHRSPSERRHSPSEKSHHSPERSHHSPSERRHS 763
 ++P+E + P+E + +P+E +P+E ++P+EK+ ++P+E + ++P+E +
 Sbjct: 413 TTYAPTEETPYEPTTEETTYTTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYE 472
 Query: 764 PLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPE 822
 P E + ++ + + +P+E ++ S E + + E +++P+E++ P E + +P+E
 Sbjct: 473 PTEETTYAPTKEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEE 532
 Query: 823 RGHSSSGKCHSPSERSHRSPSGMRQGRTERSRRSCERTRHSPSEMRPGRPSGRNHCS 892
 ++ + +T ++P+E + +P+ +E + E T ++P+E P+ +
 Sbjct: 533 TTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYA 592
 Query: 883 PSERSRRSPLKEGLKYSFPGER 905
 P E + +P +E Y+ E P
 Sbjct: 593 PIEETTYAPTEE-TTYAPAEETP 614

Score = 151 (22.7 bits), Expect = 2.0e-06, P = 2.0e-06
 Identities = 45/198 (22%), Positives = 103/198 (52%)

Query: 716 PSERSHCSPSERRHRSPSERRHSPSEKSHHSPERSHHSPSERRHRSPLERSRHSLLER 775
 PS+ + +P+E P E +PSE + ++P E + ++P+E+ +E + + + E
 Sbjct: 303 PSDETE-APTEGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPYD--VEETTY-VTEE 358
 Query: 776 SHRSPSERRSHRSFERSHRRISERS-----HSPSEKSHLSPLERSRCSPEERRGHSSS 828
 S +P++ ++ ER H E+ ++P+E++ +P E + +P+E ++ +
 Sbjct: 359 STYAPTKESETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPT 418
 Query: 829 GKCHSPSERSHRSPSGMRQGRTERSRRSCERTRHSPSEMRPGRPSGRNHCSPSERSR 888
 +T + P+E + +P+ +E + + E+T ++P+E P+ P+E +
 Sbjct: 419 EETPYEPTTEETTYTTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTTEET 478
 Query: 889 RSPLKEGLKYSFPGERPSHLSRD 912
 +P KE Y+ P E +++ + +
 Sbjct: 479 YAPTKE-TTYA-PTEETTYASTE 500

Pedant information for DKFZphtes3_8g11, frame 2

Report for DKFZphtes3_8g11.2

[LENGTH] 954
 [MW] 110063.05
 [pI] 11.40
 [PROSITE] ATP_GTP_A 1
 [KW] Irregular
 [KW] LOW_COMPLEXITY 27.67 %

SEQ ESSLSIFYDREDLVPMEESQSDSQTRISESQHSLKPNYLSQAKTDFSEQFQLLEDLQ
 SEGXXXXXXXXXXXX.....
 PRD ccccccXX
 SEQ LKIAAKLLRSQIPDPVPPPLASGLVLKYPICLQCGRCGLNCHHLQTTSGPYLLIYPQL
 SEG
 PRD hhhhhhhhhhhcc
 SEQ HLVRTPEGHGEVRLHLGFRRLRIGKRSQISKYRERDRPVIRRSPIPSQKAKIYTQASKS
 SEG
 PRD hcc
 SEQ PTSTIDLQSGSPQSPAPVQYIIRRGQSRPDLVEKTKTRAPGHYEFTQVHNLPESDSEST

Prosites for DKFZphtes3_8g11.2

(No Pfam data available for DKFZphtes3_8g11.2)

DKF2phtes3_8g5

group: testes derived

DKF2phtes3_8g5 encodes a novel 544 amino acid protein nearly identical to human KIAA087 protein.

The novel protein is a new splice variant of KIAA087.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

KIAA087, alternative spliced

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp

No poly A stretch found, no polyadenylation signal found

```

1 CCGACATCGG CCGTGTCTCC AGCACCTGCC GCGCGCTGCG CGAGCTGTGC
51 CAGAGCAGCG GGAAGGTGTG GAAGGAGCAG TTCCGGGTGA GGTGACCTTC
101 CCTTATGAAA CACTACAGCC CCACCGACTA CGTCAATTGG TTGGAAGAGT
151 ATAAAGTTCC GCAAAAAGCT GGGTTAGAAG CGCGGAAGAT TGAGCCTCG
201 TTCTCAAAGA GGTTCCTTTC AGAGCAGGTT CCTTGTAATG GCTTCAGTGA
251 CATTGAGAAC CTGGAAGGAC CAGAGATTTT TTTGAGGAT GAACTGGTGT
301 GTATCCTAAA TATGGAAGGA AGAAAAGCTT TGACCTGGAA ATACTACGCA
351 AAAAAAATTC TTACTACCT GCGGCAACAG AAGATCTTAA ATAATCTTAA
401 GGCCTTCTCT CAGCAGCCAG ATGACTATGA GTCGTATCTT GAAGGTGCTG
451 TATATATTGA CCAGTACTGC AATCCTCTCT CCGACATCAG CCTCAAAGAC
501 ATCCAGGCCC AAATTGACAG CATCGTGGAG CTTGTTTGCA AAACCTTTCG
551 GGGCATAAAC AGTCGCCACC CCAGCTTGGC CTTCAAGGCA GGTGAATCAT
601 CCATGATAAT GGAAATAGAA CTCCAGAGCC AGGTGCTGGA TGCCATGAAC
651 TATGTCCTTT ACGACCAACT GAAGTTCAAG GGGAAATCGAA TGGATTACTA
701 TAATGCCCTC AACTTATATA TGCATCAGGT TTTGATTTCG AGAACAGGAA
751 TCCCAATCAG CATGTCTCTG CTCTATTGTA CAATTGCTCG GCAGTTGGGA
801 GTCCCACTGG AGCCTGTCAA CTTCCTCAAGT CACTTCTTAT TAAGGTGGTG
851 CCAAGGCGCA GAAGGGGCGA CCCTGGACAT CTTTGACTAC ATCTACATAG
901 ATGCTTTTGG GAAAGGCAAG CAGCTGACAG TGAAGAATG CGAGTACTTG
951 ATCGGCCGAC ACGTGACTGC AGCACTGTAT GGGGTGGTCA ATGTCAAGAA
1001 GGTGTTACAG AGAATGGTGG GAAACCTGTT AAGCCTGGGG AAGCGGGAAG
1051 GCATCGACCA GTCATACCAG CTCCTGAGAG ACTCGTGGGA TCTCTATCTG
1101 GCAATGTACC CGGACCAGGT GCAGCTTCTC CTCCTCCAAG CCAGGCTTTA
1151 CTTCACCTTG GGAATCTGGC CAGAGAAGTC TTTCTGTCTT GTTTTGAAGG
1201 TGCTTGACAT CCTCCAGCAC ATCCAAACCC TAGACCCGGG GCAGCAGCGG
1251 GCGGTGGGCT ACCTGGTGCA GCACACTCTA GAGCACATTG AGCGCAAAAA
1301 GGAGGAGGTG GCGGTAGAGG TGAAGCTGCG CTCGATGAG AAGCACAGAG
1351 ATGTCTGTGA CTCCATCGGG CTCAATTATG AGCATAAGAG GTATGGCTAT
1401 AACTGTGTGA TCTACGGCTG GGACCCACCC TGCATGATGG GACACGAGTG
1451 GATCCGGAAC ATGAACGTCC ACAGCCTGCC GCACGGCCAC CACCAGCCTT
1501 TCTATAACGT GCTGGTGGAG GACGGCTCCT GTCGATACGC AGCCCAAGAA
1551 AACTTGGAAT ATAACGTGGA GCCTCAAGAA ATCTCACACC CTGACGTGGG
1601 ACGCTATTTT TCAGAGTTTA CTGGCACTCA CTACATCCCA AACGCAGAGC
1651 TGGAGATCCG GTATCCAGAA GATCTGGAGT TTGTCTATGA AACGGTGCAG
1701 AATATTTACA GTGCAAGAA AGAGAACATA GATGAGTAAA GTCTAGAGAG
1751 GACATTGCAC CTTTGCTGCT GCTGCTATCT TCCAAGAGAA CGGGACTCCG
1801 GAAGAAGACG TCTCCACGGA GCCCTCGGGA CCTGCTGCAC CAGGAAAGCC
1851 ACTCCACCAG TAGTGCTGGT TGCCTCCTAC TAAGTTTAAA TACCGTGTGC
1901 TCTTCCCAG CTGCAAGAC AATGTTGCTC TCCGCCATA CTAGTGAATT
1951 AATCTGAAAG GCACTGTGTC AGTGGCATGG CTTGTATGCT TGTCTGTGG
2001 TGACAGTTTG TGACATTCTG TCTTCATGAG GTCTCACAGT CGACGCTCCT
2051 GTAATCATTC TTTGTATTCA CTCCATTCCT CTGTCTGTCT GCATTTGTCT
2101 CAGAACATTT CTTGGCTGG ACAGATGGGG TTATGCATTT GCAATAATTT
2151 CCTTCTGATT TCTCTGTGGA ACGTGTTCGG TCCCGAGTGA GGACTGTGTG
2201 TCTTTTTACC CTGAAGTTAG TTGCATATTC AGAGGTAAAG TTGTGTGCTA
2251 TCTTGGCAGC ATCTTAGAGA TGGAGACATT AACAAGCTAA TGGTAATTAG
2301 AATCATTTGA ATTTATTTTT TTCTAATATG TGAACACAG ATTTCAAGTG
2351 TTTTATCTTT TTTTTTTTTA AATTAAATG GGAATATAAC ACAGTTTTC
2401 CTTCATATT CCTCTCTTGA GTTTATGCAC ATCTCTATAA ATCATTTAGT
2451 TTCTATTTTA TTACATAAAA TTCTTTTAGA AAATGCAAA AGTGAACTTT
2501 GTGAATGGAT TTTTCCATAC TCATCTACAA TTCCTCCATT TTAATGACT
2551 ACTTTTATTT TTTAATTAA AAAATCTACT TCAGTATCAT GAGTAGGTCT
2601 TACATCAGTG ATGGGTCTTT TTTGTAGTGA GACATACAAA TCTGATGTTA

```

2651 ATGTTTGCTC TTAGAAGTCA TACTCCATGG TCTTCAAAGA CCAAAAAATG
 2701 AGGTTTGTCT TTTGTAATCA GGAAAAAATA AATTAATGAA CCTTAAAAAA
 2751 AAAAAAATAA GG

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 105 bp to 1736 bp; peptide length: 544
 Category: known protein
 Classification: unclassified

1 MKHYSPTDYV NWLEEKVRQ KAGLEARKIV ASFSKRFFSE HVPCNGFSDI
 51 ENLEGPEIFF EDELVCILNM EGRKALTWKY YAKKILYYLR QOKILNNLKA
 101 FLQPPDDYES YLEGAVYIDQ YCNPLSDISL KDIQAQIDSI VELVCKTLRG
 151 INSRHPSLAF KAGESSMIME IELQSQVLDA MNYVLYDQK FKGNRMDYYN
 201 ALNLYMHQVL IRRGTGIPISM SLLYLTIRQ LGVPLEPVNF PSHELLRWCO
 251 GAEGATLDIF DYIYIDAFGK GKQLTVKECE YLIGQHVTAAL YGVVNVKKV
 301 LQRMVGNLLS LGKREGIDQS YQLLRDSL DL YLAMYPDQVQ LLLQARLYF
 351 HLGWPEKSF CLVLKVL DIL QHIQTLDPGQ HGAVGYLVQH TLEHIERKKE
 401 EVGVEVKLR DEKHRDVCYS IGLIMKHRY GYNCVIYWD PTCMMGHEWI
 451 RNMNVHSLPH GHHPFYNVL VEDGSCRYAA QENLEYNVEP QEISHPDVGR
 501 YFSEFTGTHY IPNAELEIRY PEDLEFVYET VQNIYSAPKE NIDE

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8g5, frame 3

TREMBLNEW:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein";
 Homo sapiens mRNA for KIAA0875 protein, partial cds., N = 1, Score =
 2832, P = 5.5e-295

>TREMBLNEW:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein"; Homo
 sapiens mRNA for KIAA0875 protein, partial cds.
 Length = 621

HSPs:

Score = 2832 (424.9 bits), Expect = 5.5e-295, P = 5.5e-295
 Identities = 537/544 (98%), Positives = 537/544 (98%)

Query: 1 MKHYSPTDYVNWLEEKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF 60
 MKHYSPTDYVNWLEEKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF
 Sbjct: 85 MKHYSPTDYVNWLEEKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF 144

Query: 61 EDELVCILNMEGRKALTWKYYAKKILYYLRQOKILNNLKAFLQPPDDYESYLEGAVYIDQ 120
 EDELVCILNMEGRKALTWKYYAKKILYYLRQOKILNNLKAFLQPPDDYESYLEGAVYIDQ
 Sbjct: 145 EDELVCILNMEGRKALTWKYYAKKILYYLRQOKILNNLKAFLQPPDDYESYLEGAVYIDQ 204

Query: 121 YCNPLSDISLKDIAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA 180
 YCNPLSDISLKDIAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA
 Sbjct: 205 YCNPLSDISLKDIAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA 264

Query: 181 MNYVLYDQKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIRQLGVPLEPVNF 240
 MNYVLYDQKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIRQLGVPLEPVNF
 Sbjct: 265 MNYVLYDQKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIRQLGVPLEPVNF 324

Query: 241 PSHELLRWCCGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV 300
 PSHELLRWCCGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV
 Sbjct: 325 PSHELLRWCCGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV 384

Query: 301 LQRMVGNLLSLGKREGIDQSYQLLRDSL DLYLAMYPDQVQ LLLQARLYFHLGIWPEKSF 360
 LQRMVGNLLSLGKREGIDQSYQLLRDSL DLYLAMYPDQVQ LLLQARLYFHLGIWPEK

979

DKFZphtes3_8m10

group: nucleic acid management

DKFZphtes3_8m10 encodes a novel 221 amino acid protein with strong similarity to polyadenylate-binding proteins.

The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.

The new protein can find application in modulation of mRNA translation and processing/stability.

strong similarity to polyadenylate-binding protein

frame shift at Bp 707-710

Sequenced by MediGenomix

Locus: unknown

Insert length: 2107 bp

Poly A stretch at pos. 2052, polyadenylation signal at pos. 2033

```
1 CGGAAAGGTC GCGGCTTGTC TGCCTGCGGG CAGCCGTGCC GAGAATGAAC
51 CCCAGCACCC CCAGCTACCC AACGGCCTCG CTCTACGTGG GGGACCTCCA
101 CCCGACGCTG ACTGAGGCGA TGCTCTACGA GAAGTTCAGC CCGGCAGGGC
151 CCATCCTCTC CATCCGGATC TGCAGGGACT TGATCACCAG CGGCTCCTCC
201 AACTACGCGT ATGTGAACCT CCAGCATACG AAGGACGCGG AGCATGCTCT
251 GGACACCATG AATTTTGATG TTATAAAGGG CAAGCCAGTA CGCATCATGT
301 GGTCTCAGCG TGATCCATCA CTTGCAAAAA GTGGAGTGGG CAACATATTC
351 GTTAAAAATC TGGATAAGTC CATTAAATAA AAAGCACTGT ATGATACAGT
401 TTCTGCTTTT GGTAAACATCC TTTCGTGTAA CGTGGTTTGT GATGAAAAATG
451 GTTCCAAGGG TTATGGATTG GTACACTTTG AGACACACGA AGCAGCTGAA
501 AGAGCTATTA AAAAAATGAA CGGAATGCTC CTAAATGGTC GCAAAGTATT
551 TGTGGACAAA TTTAAGTCTC GTAAAGAACG AGAAGCTGAA CTTGGAGCTA
601 GGGCAAAAGA GTTCCCAAT GTTTACATCA AGAATTTTGG AGAAGACATG
651 GATGATGAGC GCCTTAAGGA TCTCTTTGGC AAGTTCGGGC CCGCCTTAAG
701 TGTGAATTAA TGACCGATGA AAGTGGAAAA TCCAAAGGAT TTGGATTTGT
751 AAGCTTTGAA AGGCATGAAG ATGCACAGAA AGCTGTAGAT GAGATGAATG
801 GAAAGGAGCT CAATGGAAAA CAAATTTACG TTGGTCGAGC TCAGAAAAAA
851 GTGGAACGGC AGACGGAACT TAAGCGCACA TTTGAACAGA TGAAGCAAGA
901 TAGGATCACC AGATACCAGG TTGTTAATCT TTATGTGAAA AATCTTGATG
951 ATGGTATTGA TGATGAACGT CTCCGGAAG CGTTTCTCC ATTTGGTACA
1001 ATCAGTAGTG CAAAGGTTAT GATGGAAGGT GGTGCGAGCA AAGGGTTTGG
1051 TTTTGTATGT TTCTCCTCCC CAGAAGAAGC CACTAAAGCA GTTACAGAAA
1101 TGAACGGTAG AATTGTGGCC ACAAAGCCAT TGTATGTAGC TTTAGCTCAG
1151 CGCAAAGAAG AGCGCCAGGC TTACCTCACT AACGAGTATA TGCAGAGAAT
1201 GGCAAGTGTA CGAGCTGTGC CCAACCAGCG AGCACCTCCT TCAGGTTACT
1251 TCATGACAGC TGTCCACAG ACTCAGAACC ATGCTGCATA CTATCCTCCT
1301 AGCCAAATTG CTCGACTAAG ACCAAGTCCT CGCTGGACTG CTCAGGGTGC
1351 CAGACCTCAT CCATTCCAAA ATAAGCCCAG TGCTATCCGC CCAGGTGCTC
1401 CTAGAGTACC ATTTAGTACT ATGAGACCAG CTTCCTCACA GGTTCACAGA
1451 GTCATGTCAA CGCAGCGTGT TGCTAACACA TCAACACAGA CAGTGGGTCC
1501 ACGTCTTGCA GCTGCTGCTG CTGCTGCAGC TACCCCTGCT GTGCGCACGG
1551 TTCCACGGTA TAAATATGCT GCGGGAGTTC GCAATCCTCA GCAACATCGT
1601 AATGCACAGC CACAAGTTAC AATGCAACAG CTGCTGTTC ATGTACAAGG
1651 TCAGGAAACT TTGACTGCCT CCAGGTTGGC ATCTGCCCTT CCTCAAAAGC
1701 AAAAGCAAAT GTTAGGTGAA CGGCTCTTTC CTCTTATCA AGCCATGCAC
1751 CCTACTCTTG CTGGGAAAAT CACTGGCATG TTGTTGGAGA TTGATAATTC
1801 AGAAGTCTTT TATATGCTCG AGTCTCCAGA GTCACTCCGT TCTAAGGTTG
1851 ATGAAGCTGT AGCTGTACTA CAAGCCCACC AAGCTAAAGA GGCTACCCAG
1901 AAAGCAGTTA ACAGTGCTAC CGGTGTTCCA ACTGTTTAAA ATTGATCAGA
1951 GACCCAGAAA AGAAATTTGT GCTTCACCGA AGAAAAATAT CTAACATCG
2001 AGAAACTATG GGAACAAAAA TTGCAAAATC TAAATAAAAA AATGCAAAAT
2051 CTAATAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
2101 AAAAAGG
```

BLAST Results

Entry HSPOLYAB from database EMBL:
Human mRNA for polyA binding protein
Score = 5420, P = 0.0e+00, identities = 1162/1243

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 707 bp to 1936 bp; peptide length: 410
 Category: strong similarity to known protein
 Classification: unset
 Prosite motifs: RNP_1 (10-18)
 RNP_1 (112-120)

```

1 LMTDESGKSK GFGFVSFERH EDAQKAVDEM NGKELNGKQI YVGRAQKKVE
51 RQTELKRTFE QMKQDRITRY QVVNLYVKNL DDGIDDERLR KAFSPFGTIT
101 SAKVMEGGR SKGFGFVCF S PEEATKAVT EMNGRIVATK PLYVALAQRK
151 EERQAYLTNE YMQRMASVRA VPNQRAPPSG YFMTAVPQTQ NHAAYYPPSQ
201 IARLRPSPRW TAQGARPHPF QNKPSAIRPG APRVFFSTMR PASSQVPRVM
251 STQRVANTST QTVGPRPAAA AAAAATPAVR TVPRYKYAAG VRNPQQHRNA
301 QPQVTMQOLA VHVQGQETLT ASRLASAPPQ KQKQMLGERL FPLIQAMHPT
351 LAGKITGMLL EIDNSELLEYM LESPELRSK VDEAVAVLQA HQAKEATQKA
401 VNSATGVPTV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8ml0, frame 2

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1931,
 P = 1.7e-199

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1928, P =
 3.6e-199

>PIR:DNHUPA polyadenylate-binding protein - human
 Length = 633

HSPs:

Score = 1931 (289.7 bits), Expect = 1.7e-199, P = 1.7e-199
 Identities = 384/415 (92%), Positives = 394/415 (94%)

```

Query:      1 LMTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFE 60
            +MTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKR FE
Sbjct:     219 VMTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRKFE 278

Query:     61 QMKQDRITRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITS AKVMEGGRSKGFGFVCF 120
            QMKQDRITRYQ VNLVKNLDDGIDDERLRK FSPFGTITS AKVMEGGRSKGFGFVCF
Sbjct:     279 QMKQDRITRYQGVNLYVKNLDDGIDDERLRKEFSPFGTITS AKVMEGGRSKGFGFVCF 338

Query:     121 SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQRMASVRAPN-----Q 174
            SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQA+LTN+YMQRMASVRAPN Q
Sbjct:     339 SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAHLTNQYMQRMASVRAPNPVINPYQ 398

Query:     175 RPPSGYFMTAVPQTQNHAAAYPPSQIARLRPSPRWTAQGARPHPFQNKPSAIRPGAPRV 234
            APPSGYEM A+PQTQN AAYYPPSQ+A+LRPSPRWTAQGARPHPFQNP AIRP APR
Sbjct:     399 PAPPSGYFMAAIPQTQNRAAYYPPSQVAQLRPSPRWTAQGARPHPFQNMPGAIRPAAPRP 458

Query:     235 PFSTMRPASSQVPRVMSTQRVANTSTQTVGPRPAAAAAATPAVRTVPRYKYAAGVRNP 294
            PFSTMRPASSQVPRVMSTQRVANTSTQT+GPRPAAAAA TPAVRTVP+YKYAAGVRNP
Sbjct:     459 PFSTMRPASSQVPRVMSTQRVANTSTQTMGPRPAAAAA-TPAVRTVPQYKYAAGVRNP 517

Query:     295 QQHRNAQPQVTMQOLAVHVQGQETLTASRLASAPPQKQKQMLGERLFPLIQAMHPTLAGK 354
            QQH NAQPQVTMQQ AVHVQGQE LTAS LASAPPQ+QKQMLGERLFPLIQAMHPTLAGK
Sbjct:     518 QQH LNAQPQVTMQQPAVHVQGQEPLTASRLASAPPQEQKQMLGERLFPLIQAMHPTLAGK 577

Query:     355 ITGMLLEIDNSELLEYMLESPELRSKVDEAVAVLQAQAKEATQKAVNSATGVPTV 410
            ITGMLLEIDNSELL+MLESPELRSKVDEAVAVLQAQAKEA QKAVNSATGVPTV
Sbjct:     578 ITGMLLEIDNSELLHMLSPESLRSKVDEAVAVLQAQAKEAAQKAVNSATGVPTV 633

```

Score = 315 (47.3 bits), Expect = 1.9e-27, P = 1.9e-27

Identities = 71/163 (43%), Positives = 102/163 (62%)

Query: 1 LMTDESGKSGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFE 60
 ++ DE+G SKG+GFV FE E A++A+++MNG LN ++++VGR + + ER+ EL +
 Sbjct: 130 VVCDENG-SKGYGFVHFETQEAARAIEKMNGMLNDRKVFVGRFKSRKEREAEELGARAK 188

Query: 61 QMKQDRITRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMM-EGGRSKGFGFVCF 119
 + N+Y+KN + +DDERL+ F P S KVM E G+SKGFGFV F
 Sbjct: 189 EF-----TNVYIKNFGEDMDDERLKDLFGP---ALSVKVMTDESGKSGFGFVSF 235

Query: 120 SSPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQ 163
 E+A KAV EMNG+ + K +YV AQ+K ERQ L ++ Q
 Sbjct: 236 ERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRKFEQ 279

Score = 214 (32.1 bits), Expect = 1.9e-14, P = 1.9e-14
 Identities = 50/150 (33%), Positives = 87/150 (58%)

Query: 8 KSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFEQMKQDRI 67
 +S G+ +V+F++ DA++A+D MN + GK + + +Q R L+++
 Sbjct: 50 RSLGYAYVNFQQPADAERALDTMNFVIGKGPVRIMWSQ----RDPSLRKS----- 96

Query: 68 TRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMMEGGRSKGFGFVCFSSPEEATK 127
 V N+++KNLD ID++ L FS FG I S KV+ + SKG+GFV F + E A +
 Sbjct: 97 ---GVGNIFIKNLDSIDNKALYDTFSAFGNILSCKVVCDENGSKGYGFVHFETQEAAR 153

Query: 128 AVTEMNGRIVATKPLYVALAQRKEERQAYL 157
 A+ +MNG ++ + ++V + ++ER+A L
 Sbjct: 154 AIEKMNGMLNDRKVFVGRFKSRKEREAEEL 183

Score = 120 (18.0 bits), Expect = 4.8e-04, P = 4.8e-04
 Identities = 30/99 (30%), Positives = 54/99 (54%)

Query: 70 YQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVM--MEGGRSKGFGFVCFSSPEEATK 127
 Y + +LYV +L + + L + FSP G I S +V M RS G+ +V F P +A +
 Sbjct: 8 YPMASLYVGDLPDVTEAMLYEKFSPAGPILSIRVCRDMITRRSLGYAYVNFQQPADAER 67

Query: 128 AVTEMNGRIVATKPLYVALAQRKEE-RQAYLTNEYMQRM 165
 A+ MN ++ KP+ + +QR R++ + N +++ +
 Sbjct: 68 ALDTMNFVIGKGPVRIMWSQRDPSLRKSGVGNIFIKNL 106

Peptide information for frame 3

ORF from 45 bp to 707 bp; peptide length: 221
 Category: strong similarity to known protein
 Classification: unset
 Prosite motifs: RNP_1 (138-146)

1 MNPSTPSYPT ASLYVGDLP DVTEAMLYEK FSPAGPILSI RICRDLITSG
 51 SSNYAYVNFQ HTKDAEHALD TMNFDVIK GK PVRIMWSQ RD PSLRKSGVGN
 101 IFVKNLDSI NNKALYDTVS AFGNILSCNV VCDENGSKGY GFVHFETHEA
 151 AERAIAKKMNG MLLNGRKVEV GQFKSRKERE AELGARAKEF PNVYIKNFGE
 201 DMDDERLKDL FGKFGPALSV N

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_8m10, frame 3

SWISSPROT:PAB1 HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
 PROTEIN 1) (PABP 1)., N = 1, Score = 1039, P = 5.7e-105

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1031, P =
 4e-104

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1009,
 P = 8.7e-102

>SWISSPROT:PAB1 HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
 PROTEIN 1) (PABP 1).
 Length = 636

HSPs:

Score = 1039 (155.9 bits), Expect = 5.7e-105, P = 5.7e-105
Identities = 199/220 (90%), Positives = 205/220 (93%)

Query: 1 MNPSTPSYPTASLYVGDLPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQ 60
MNPSPSYPTASLYVGDLPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQ 60
Sbjct: 1 MNPSPSYPTASLYVGDLPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQ 60

Query: 61 HTKDAEHALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNLDSINNKALEYDTVS 120
DAE ALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNLDSINNKALEYDTVS 120
Sbjct: 61 QPADAERALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNLDSINNKALEYDTVS 120

Query: 121 AFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFQSRKERE 180
AFGNILSC NVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFQSRKERE 180
Sbjct: 121 AFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFQSRKERE 180

Query: 181 AELGARAKEFPNVYIKNFGEDMDERLKDLPFGKFGPALS 220
AELGARAKEFP NVYIKNFGEDMDERLKDLPFGKFGPALS 220
Sbjct: 181 AELGARAKEFPNVYIKNFGEDMDERLKDLPFGKFGPALS 220

Score = 275 (41.3 bits), Expect = 4.1e-23, P = 4.1e-23
Identities = 71/233 (30%), Positives = 120/233 (51%)

Query: 2 NPSTPSYPTASLYVGDLPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQ 61
+PS +++++L + LY+ FS G ILS ++ D S + + Q
Sbjct: 90 DPSLRKSGVGNIFVKNLDSINNKALEYDTVS AFGNILSCNVVCDENGSKGYGFVHFETQ 149

Query: 62 TKD-AEHALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNLDSINNKALEYDTVS 117
+ A ++ M + K R +R+ L R N+++KN + +++ L D
Sbjct: 150 AAERAIKKMNGMLLNGRKVFVGQFQSRKERE AELGARAKEFPNVYIKNFGEDMDERLKD 209

Query: 118 TVSAFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFQSR 176
FG LS V+ DE+G SKG+GFV FE HE A++A+ +MNG LNG+++VG+ + +
Sbjct: 210 LFGKFGPALS VKVMTDESGKSGFGFVS FERHEDAQKAVDEMNGKELNGKQIYVGRQKK 269

Query: 177 KEREAEELGARAKEFP-----NVYIKNFGEDMDERLKDLPFGKFGPALS 219
ER+ EL + ++ N+Y+KN + +DDERL+ F FG S
Sbjct: 270 VERQTELKRRKFEQMKQDRITRYQGVNLYVKNLDDGIDDERLRKEFS PFGTITS 322

Score = 227 (34.1 bits), Expect = 6.3e-18, P = 6.3e-18
Identities = 57/187 (30%), Positives = 101/187 (54%)

Query: 12 SLYVGDLPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQHTKDAEHALDT 71
++Y+ + D+ + L+ F GP LS+++ D + S + +V+F+ +DA+ A+D
Sbjct: 192 NVYIKNFGEDMDERLKDLPFGKFGPALS VKVMTDESGKSGFGFVS FERHEDAQKAVDE 250

Query: 72 MNFVDVIKGPVRIMWSQR-----DPSLRKSGVGNIFVKNLDSINNKA 114
MN + GK + + +Q+ D R GV N++VKNLD I+++
Sbjct: 251 MNGKELNGKQIYVGRQKKVERQTELKRRKFEQMKQDRITRYQGV-NLYVKNLDDGIDDER 309

Query: 115 LYDTVS AFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFQ 174
L S FG I S V+ + SKG+GFV F + E A +A+ +MNG ++ + +V +
Sbjct: 310 LRKEFS PFGTITS AKVMMEGGRSKGFGFVCFSSPEEATKAVTEMNGRIVATKPLYVALAQ 369

Query: 175 SRKEREAEEL 183
++ER+A L
Sbjct: 370 RKEERQAH 378

Score = 100 (15.0 bits), Expect = 2.3e-02, P = 2.3e-02
Identities = 26/99 (26%), Positives = 53/99 (53%)

Query: 8 YPTASLYVGDLPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQHTKDAE 66
Y +LYV +L + + L ++FSP G I S ++ ++ G S + +V F ++A
Sbjct: 291 YQGVNLYVKNLDDGIDDERLRKEFS PFGTITS AKV---MMEGGRSKGFGFVCFSSPEEAT 347

Query: 67 HALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNL 106
A+ MN ++ KP+ + +QR R++ + N +++ +
Sbjct: 348 KAVTEMNGRIVATKPLYVALAQKREE-RQAHLTNQYMQRM 386

Pedant information for DKFZphtes3_8m10, frame 2

Report for DKFZphtes3_8m10.2

[LENGTH] 409
[MW] 45235.68
[pI] 10.08
[HOMOL] SWISSPROT: PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN
1) (PABP 1). 0.0

[FUNCAT] 04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S. cerevisiae, YER165w] 1e-54

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1e-54

[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YER165w] 1e-54

[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YER165w] 1e-54

[FUNCAT] 04.05.99 other mrna-transcription activities [S. cerevisiae, YNL016w] 1e-15

[FUNCAT] 11.01 stress response [S. cerevisiae, YGR159c] 1e-12

[FUNCAT] 04.01.04 rna processing [S. cerevisiae, YGR159c] 1e-12

[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YNL175c] 4e-09

[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YPR112c] 5e-08

[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YHR086w] 3e-07

[FUNCAT] 03.13 meiosis [S. cerevisiae, YHR086w] 3e-07

[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YHR086w] 3e-07

[FUNCAT] 04.07 rna transport [S. cerevisiae, YOL123w HRP1 - CF Ib] 9e-07

[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YCL011c] 3e-06

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGR250c] 8e-06

[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR432w] 2e-05

[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDR432w] 2e-05

[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YFR023w] 3e-05

[FUNCAT] 03.01 cell growth [S. cerevisiae, YBR212w] 3e-04

[BLOCKS] BL00030B Eukaryotic RNA-binding region RNP-1 proteins

[SCOP] dlsx1_4.34.7.1.3 Sex-lethal protein [(Drosophila melanogaster) 1e-17

[PIRKW] nucleus 0.0

[PIRKW] duplication 0.0

[PIRKW] RNA binding 0.0

[PIRKW] nucleolus 2e-09

[PIRKW] tandem repeat 2e-09

[PIRKW] single-stranded DNA binding 3e-06

[PIRKW] DNA binding 5e-13

[PIRKW] phosphoprotein 6e-10

[PIRKW] ribosome 3e-08

[PIRKW] mitochondrion 3e-08

[PIRKW] alternative splicing 9e-11

[PIRKW] chloroplast 2e-19

[PIRKW] transcription regulation 2e-07

[PIRKW] protein biosynthesis 3e-08

[SUPFAM] nucleolin 6e-10

[SUPFAM] glycine-rich RNA-binding protein 2e-07

[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 2e-19

[SUPFAM] polyadenylate-binding protein 0.0

[SUPFAM] ribonucleoprotein repeat homology 0.0

[PROSITE] RNP_1 2

[PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)

[KW] Irregular

[KW] 3D

[KW] LOW_COMPLEXITY 5.62 %

SEQ MTDES GSKSGFGFVS FERHEDA QKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFEQ

SEG

1sxl-

SEQ MKQDRITRYQVNVLYVKNLDDGIDDERLRKAFSPFGTITS AKVMMEGGRSKGFGFVCFSS

SEG

1sxl-CEEEECCTTTTHHHHHHHHTTTTCCCCCEECTTTCTTTEEEECTTT

SEQ PEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQRMASVRAVPNQRPAPPSGY

SEG

1sxl- HHHHHHHHHHTTTCCCCCBBCCBCC.....

SEQ FMTAVPQTQNHAAYPPSQIARLRPSRWTAQGARPFPFQNKPSAIRPGAPRVPFSTMRP

SEG

1sxl-

SEQ ASSQVPRVMSTQRVANTSTQTVGPRPAAAAAATPAVRTVPYKYAAGVRNPQQHRNAQ

SEGXXXXXXXXXXXXXXXXXXXXX.....

1sxl-

SEQ PQVTMQQLAVHVQGQETLTASRLASAPPQKQKQLGERLFPLIQAMHPTLAGKITGMLE

SEG

1sxl-

SEQ IDNSELMLSPESLRSKVDEAVAVLQAHQAKEATQKAVNSATGVPTV

SEG

1sxl-

Prosites for DKFZphtes3_8m10.2

PS00030	9->17	RNP_1	PD0000030
PS00030	111->119	RNP_1	PD0000030

Pfam for DKFZphtes3_8m10.2

HMM_NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)		
HMM	*IYVGNLPWDtTEEDLrDlFsQFGpIvsIrmMrDRetGRSRGFVFEFED +YV+NL+ +++E LR +FS+FG I+S+++M+ E GRS+GF+V F +		
Query	74	LYVKNLDDGIDDERLRKAFSPFGTITSKVMM--EGGRSKGFGFVCFSS	120
HMM	EEDAekAIdemNGmeFmGRrIRV* +E+A+KA+ EMNG+++ ++++V		
Query	121	PEEATKAVTEMNGRIVATKPLYV	143

Pedant information for DKFZphtes3_8m10, frame 3

Report for DKFZphtes3_8m10.3

```

[LENGTH]      235
[MW]           26308.08
[pI]           8.95
[HOMOL]        SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN
1) (PABP 1). 1e-113
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YER165w] 1e-64
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1e-64
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YER165w] 1e-64
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YER165w] 1e-64
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YFR023w] 1e-24
[FUNCAT]       11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YFR023w] 1e-24
[FUNCAT]       04.05.99 other mrna-transcription activities [S. cerevisiae, YNL016w]
2e-19
[FUNCAT]       04.05.03 mrna processing (splicing) [S. cerevisiae, YOR319w] 2e-14
[FUNCAT]       04.01.04 rrna processing [S. cerevisiae, YGR159c] 1e-11
[FUNCAT]       11.01 stress response [S. cerevisiae, YGR159c] 1e-11
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGR250c] 1e-09
[FUNCAT]       04.07 rna transport [S. cerevisiae, YOL123w HRP1 - CF Ib] 1e-09
[FUNCAT]       30.13 organization of chromosome structure [S. cerevisiae, YCL011c] 8e-09
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YPR112c] 2e-08
[FUNCAT]       03.13 meiosis [S. cerevisiae, YHR086w] 2e-08
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YBR212w] 3e-08
[FUNCAT]       03.01 cell growth [S. cerevisiae, YBR212w] 3e-08
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR432w]
3e-04
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YDR432w] 3e-04
[BLOCKS]       BL00030B Eukaryotic RNA-binding region RNP-1 proteins
[BLOCKS]       BL00900D Bacteriophage-type RNA polymerase family proteins signatur
[SCOP]         disx1_ 4.34.7.1.3 Sex-lethal protein [(Drosophila melanogaster) 9e-23
[SCOP]         d2ula_ 4.34.7.1.2 U1A protein [human (Homo sapiens) 6e-24
[SCOP]         dlup1_2 4.34.7.1.1 Nuclear ribonucleoprotein A1, RNP A1, UP 1e-13
[PIRKW]        nucleus 1e-110
[PIRKW]        duplication 1e-110
[PIRKW]        RNA binding 1e-110
[PIRKW]        nucleolus 4e-10
[PIRKW]        tandem repeat 4e-10
[PIRKW]        single-stranded DNA binding 1e-06
[PIRKW]        DNA binding 9e-12
[PIRKW]        phosphoprotein 4e-10
[PIRKW]        mitochondrion 6e-07
[PIRKW]        heterotrimer 4e-06
[PIRKW]        alternative splicing 1e-15
[PIRKW]        chloroplast 5e-11
[PIRKW]        transcription regulation 3e-09
[PIRKW]        GTP binding 2e-06
[SUPFAM]       helix-destabilizing protein 1e-07
[SUPFAM]       nucleolin 4e-10
[SUPFAM]       glycine-rich RNA-binding protein 2e-07
[SUPFAM]       yeast HRP1 protein 2e-08

```

[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 3e-25
 [SUPFAM] polyadenylate-binding protein 1e-112
 [SUPFAM] ribonucleoprotein repeat homology 1e-112
 [PROSITE] RNP_1 1
 [PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)
 [KW] All_Beta
 [KW] 3D

SEQ ERSRLVCLRAAVPRMNPSTPSYPTASLYVGDLPDVTEAMLYEKFSAPGILSIRICRDL
 lhal-EEEETTTTTCHHHHHHHHGGGCCEEEEEEETT

SEQ ITSGSSNYAYVNFQHTKDAEHALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNL
 lhal- TTTCEEEEEEEECCHHHHHHHHTTEEE-TT---EEEEEECTTTTCCCCCEEEBEC

SEQ DKSINNKALYDTVSAFGNLSNVVCDENGSKGYGVHFETHAAERAIKKMNGMLNGR
 lhal- TTTTCHHHHHHHHGGGCCEEEEEEETTTCCEEEEEECCHHHHHHHH.....

SEQ KVFVGQFKSRKEREALGARAKEFPNVYIKNFGEDMDDERLKDLPFGKFGPALSVN
 lhal-

Prosite for DKFZphtes3_8m10.3

PS00030 152->160 RNP_1 PDOC00030

Pfam for DKFZphtes3_8m10.3

HMM_NAME RNA recognition motif. (aka RRM, RBD, or RNP domain)

HMM *IYVGNLPWDtTEEDLrDlFsQFGpivsIrMMrDreTGRSRGFVFEFED
 +YVG+L +D+TE +L + FS+ GPI+SIR+ RD T S +A+V+F+

Query 27 LYVGDLPDVTEAMLYEKFSAPGILSIRICRDLITSGSSNYAYVNFQH 75

HMM EEDAekAIdemNGmeFmGRrIRV*

DAE A+D+MN ++ G+++R+
 Query 76 TKDAEHALDTMNFVDVIKGPVRI 98

HMM *IYVGNLPWDtTEEDLrDlFsQFGpivsIrMMrDreTGRSRGFVFEFED
 I+V+NL+ +++ L D S FG I+S++++ D + S+G++EV FE+

Query 115 IFVKNLDSINNKALYDTVSAFGNLSNVVCD--ENGSKGYGVHFET 161

HMM EEDAekAIdemNGmeFmGRrIRV*

+E+AE+AI +MNGM+++GR++ V
 Query 162 HEAAERAIKKMNGMLNGRKFV 184

DKFZphtes3_8p7

group: testes derived

DKFZphtes3_8p7 encodes a novel 412 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits (both from testis libraries)

Sequenced by MediGenomix

Locus: unknown

Insert length: 2899 bp

Poly A stretch at pos. 2870, polyadenylation signal at pos. 2852

```
1 CCGACCCGCC CTGGGGTGCT GCGTGCGCTG CCTGCTCCCG CCTGAGGAAA
51 ACACCTGCCA TGGCGCAAGG CCGGGAGCGC GACGAAGGCC CCCACTCCGC
101 CGGCGGCGCG TCCTTGTCGG TGAGATGGGT GCAAGGATTC CTAAGCAGA
151 ATGTTTCATTT GTCAACGACA ACACCATTTG CTACCCTTGT GGAATTATG
201 TAATATTTAT TAATATTGAA ACCAAGAAAA AGACTGTACT GCAGTGTAGT
251 AATGGAATTG TGGGCGTCAT GGCAACTAAC ATCCCTGTG AAGTTGTGGC
301 TTTTCTGAC CGGAAGCTAA AACCTCTCAT CTACGTATAC AGCTTCCAG
351 GATTGACCAG AAGGACCAAA TTGAAAGGCA ACATTCTCCT GGACTACACT
401 TTACTTTCAT TCAGTTACTG TGGCACCTAC CTGGCTAGTT ACTCCTCTCT
451 CCCAGAATTT GAACTGGCCC TTTGGAATG GGAATCGAGT ATCATTGTGT
501 GTAAGAAATC ACAGCCTGGA ATGGATGTGA ACCAAATGTC TTTTAACCCC
551 ATGAACCTGGC GCCAGCTGTG CTTATCAAGT CCAAGTACAG TGAGCGTGTG
601 GACCATTGAA AGAAGTAACC AGGAGCATTT TTTAGAGCA AGGTCGGTGA
651 AATTACCTCT AGAAGATGGG TCATTTTTTA ATGAACGGA TGTCGTTTTT
701 CCCAGTCGTG TGCCGAAAGA TCTCATCTAT GGTCCCGTGC TGCCACTGTC
751 AGCCATTGCC GGGCTGGTAG GCAAGAGGCG AGAGACTTTC CGGCCGAAAG
801 ATGATCTATA TCCTTTGCTT CACCCGACTA TGCATTGCTG GACTCCAACA
851 AGTGACTTGT ACATTGGCTG TGAAGAGGGT CATCTTTTAA TGATTAATGG
901 AGACACCTTG CAAGTGACTG TACTTAATAA GATAGAAGAG GAATCGCCAT
951 TGGAAGACAG AAGAAATTTT ATCAGTCCAG TAACCTTGGT ATATCAGAAG
1001 GAGGGCGTGC TGGCTTCTGG AATTGATGGC TTTGTGTATT CTTTATTAT
1051 TAAAGATAGA AGTTACATGA TCGAGGATTT TCTTGAGATT GAAAGACCTG
1101 TAGAACATAT GACATTTTCT CCCAATTATA CAGTGTGTGCT GATTCAAACA
1151 GACAAGGGAT CTGTTTATAT CTACACTTTT GGTAAAGGAGC CAACCTTAAA
1201 TAAAGTCTTA GATGCTTGTG ATGGGAAATT TCAGGCAATT GACTTTATCA
1251 CACCTGGAAC CCAATACTTC ATGACACTTA CATATTCAGG GGAATTTTGT
1301 GTTTGGTGGC TGGAGGATTG TGCTTGTGTA AGCAAGATTT ATCTGAATAC
1351 CCTAGCAACG GTTCTGGCTT GCTGTCCATC CTCCTCTCTC GCAGCCGTGG
1401 GCACGGAGGA TGGCTCGGTC TACTTCATCA GCGTATATGA TAAGGAATCC
1451 CCTCAGTTCG TGCACAAGGC CTTTCTCTCG GAATCGTCCG TGCAGCACGT
1501 CGTGTAAGTC CTTTCTGCCT CCAGGAGCGG CTCCTGTGTA CACCCGTCTG
1551 TTGAAAATTC TAGTGAAGCC ATCCTTTCTT TTAATTTTAA GTTTTACGTG
1601 TTTTATTGTT TTTGAATGTT AATATATTCA CACAGTTCAA CACTCAAAAG
1651 GTACAGAGGG CTGTGTAGTA AAGTACCCCG CATACCCAGG TCTGTCCTTG
1701 CAGGCAGCCT GGTACCAATT TCTCATGTCT CTCCTGAGAT GTTTTATCCA
1751 TGAACAAGCA AAACATAATA AGCACTTCTT TTTACTTGTA TCAATGGCCA
1801 TCATGTGTGT ATAGTGTGCC AGGCATTCTT GCTGTATTAA CTCCATGAGG
1851 TAAACACTCT TGTGTCTCTT ATTTGACAGG TGAGGAAGAT AAGGCACAAG
1901 GATTTTAAAT AACTTGCTCA ATAGTACACA GATAGTGAAT GGCAATGTT
1951 GGGATTTGAA CCCAGGTAGT TGGGCTGCAG AGTCACTGCC TTTGCTCTTA
2001 AAAGGAGAAA ACTATGTACA ATGCCTCATT TCTTTTTCCT CTTAATCGTA
2051 TATCTTGGAG AATGTTTTAT ATCCACACAT AAAGACCAGC CTGATTATTT
2101 GTATAGCCAC ATAGTATTCC ATTATATGAA TATACTATCA TTTTTTAAAA
2151 ACGGTATATT AATGAACATT TAGAGTATTT CAAAACCTTT GAAGCAATAC
2201 TTTTAAGATG ATAATATAGA GACATTAGAT TTGGACTTGT AGGTGCTATC
2251 ATTTATTACTG TTTCTTTTTA ATTTATTATA TTATTAGGTA TTAATAAGAA
2301 CAGACATTTG TATTCTGCTT TACAGCTTGA GATCACTGTA GCTTGTGGCA
2351 TGTGATCCTC AAAACACCAG TCAGAAAGGT GTTATCTTCA TCCCTATTAG
2401 ACAAAATTAG GAATTCAGGG TTAGAGAGGT GAGGAAAAGC ATTGTCCAAG
2451 ATTACACATT ACACAGCTAG CACACTGAGG AGCTGGCCCT GCCACTGTGG
2501 ACTGCCCAGC TCCACCACCC TAGCTCAGTG GGAAGGATG GATAACCTCC
2551 TTCCATTTAC CCCCTGCCTT TCTGCACTGT CATTTTTTTG TGCCTTTCCT
2601 TTCTCAGATC CTCTTATTCT AATTTACATC TTCCCACTTT TTCTAATTTG
2651 ATAAAGTTGT AGACATGTTT CACTACATTC TTCCTCCAC TGCCAGGTAC
2701 CAGACACAGG GTAATGAAAT GTCACACCCA CCACTAATTT GAGAATTGCT
```

988

DKFZphtes3_9e22

group: testes derived

DKFZphtes3_9e22 encodes a novel 227 amino acid protein with weak partial similarity to Ring-finger proteins.

For the novel protein, Pfam, but not Prosite predicts a C3HC4 type RING finger motif. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to zinc finger proteins

Sequenced by DKFZ

Locus: unknown

Insert length: 1318 bp

Poly A stretch at pos. 1308, no polyadenylation signal found

```

1 GCTCCCCCGG CTTTCGGAGC CCGGGGGCGG CCTGTGGCGC GCGGAGCCCG
51 CGCCGGGACTG CGCCTCTTTG GACCTTGAGG GGAAACATGC GTTTGCCTTG
101 GATCGTTTGA AATTCTAAGT TTGGGATCCC CGCCCGCCCG CCTGCCTCTT
151 CGCCCGCCGCG GGTTCCTTTCC TTTTTCCTT TTGCTTTTTT TCCTTTTCTC
201 CCTCCGGGTC TCCTTTTGA CTCCTCCCCC CTTATGCTC GCCCAGCCCT
251 CCCCCTGCTG CTGAGAAGTG GGGGAGGGTC TCGGCCTCCA GGTTCCTGCC
301 CCACCGGGGC CCGGGCGAGC ATGGGGGGCA AGCAGAGCAC GCGGGCCCGC
351 TCCCGGGGGC CCTTCCCGGG GGTCTCCACC GATGACAGCG CCGTGGCCGC
401 GCGGGGAGGG GCGCCCATTT TCGGGCACTA CCGGACGGGC GCGGGGGCCA
451 TGGGGCTGCG CAGCCGCTCG GTCAGCTCGG TGGCAGGCAT GGGCATGGAC
501 CCCAGCACGG CCGGGGGGGT GCCCTTTGGC CTCTACACCC CCGCCTCCCG
551 GGGCACCCGG GACTCCGAGA GGGCGCCCGG CGGCGGAGGG TCTGCGTCCG
601 ACTCCACCTA TGCCCATGGC AATGGTTACC AGGAGACGGG CCGCGGTACAC
651 CATAGAGACG GGATGCTGTA CCTGGGCTCC CGAGCCTCGC TGGCGGATGC
701 TCTACCTCTG CACATCGCAC CCAGGTGGTT CAGCTCGCAT AGTGGTTTCA
751 AGTGCCCAT TTGCTCCAAG TCTGTGGCTT CTGACGAGAT GGAAATGCAC
801 TTTATAATGT GTTTGAGCAA ACCTCGCCTC TCCTACAACG ATGATGTGCT
851 GACTAAAGAC GCGGGTGAGT GTGTGATCTG CTTGGAGGAG CTGCTGCAGG
901 GGGACACGAT AGCCAGGCTG CCCTGCCTGT GCATCTATCA CAAAAGCTGC
951 ATAGACTCGT GGTTTGAAGT GAACAGATCT TGTCCGGAAC ACCCTGCGGA
1001 CTGACCTGCG GGCTTGCTTG CTGACTCCTC TCAAAGGGAC AGAGCGCCCC
1051 TGCTCCAGGG AGGAGGCTCA CCGGACCCCTG GGGCAGAGCT GAGCTTGGGA
1101 CACCCAGCGG AACAGGGCAC CCCTTCTGCA CTGACTTCCA GATCATGGTT
1151 CTCCTTCCT CCCTGAGGAC ACCAAATTGG ATGAGAGCAA GTTTAGAGA
1201 AGAATGAATC AACTGCTATC CTTCCCTCA CCCCTCAGCC CAGGAGGGAA
1251 AGGGCATTCT CTTTTTCATC TTTGAAAGGC ATTGTGGGTC TGCTTTTAAA
1301 GTGTTTACAA AAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 321 bp to 1001 bp; peptide length: 227
 Category: similarity to known protein
 Classification: unclassified

```

1 MGGKQSTAAR SRGPFPGVST DDSAVPPPGG APHFGHYRTG GGAMGLRSRS
51 VSSVAGMGMD PSTAGGVVFG LYTPASRG TG DSEAPCGGG SASDSTYAHG
101 NGYQETGGGH HRDGMVLYGS RASLADALPL HIAPRWFSSH SGFKPCPCSK
151 SVASDEMEMH FIMCLSKPRL SYNDDVLT KD AGECVICLEE LLQGDITARL

```


201 PCLCIYHKSC IDSWFEVNRSCPEHPAD

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9e22, frame 3

TREMBL:AF078823.1 product: "RING-H2 finger protein RHA2b"; Arabidopsis thaliana RING-H2 finger protein RHA2b mRNA, complete cds., N = 1, Score = 111, P = 2.8e-06

TREMBL:AF078822.1 product: "RING-H2 finger protein RHA2a"; Arabidopsis thaliana RING-H2 finger protein RHA2a mRNA, complete cds., N = 1, Score = 112, P = 6.6e-06

TREMBL:AC004138.14 gene: "T17M13.17"; Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence, complete sequence., N = 2, Score = 123, P = 1.4e-05

PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana, N = 1, Score = 142, P = 8.8e-08

>PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana
Length = 327

HSPs:

Score = 142 (21.3 bits), Expect = 8.8e-08, P = 8.8e-08
Identities = 24/57 (42%), Positives = 30/57 (52%)

Query: 166 SKPRLSYNDVLTGDAGECVICLEELQGDITARLPCLCIYHKSCIDSWFEVNRSCP 222
S P + LT D +C +C+EE + G LPC IYHK CI W +N SCP
Sbjct: 206 SLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATELPCKHIYHKDCIVPWLRLNNSCP 262

Pedant information for DKFZphtes3_9e22, frame 3

Report for DKFZphtes3_9e22.3

[LENGTH] 227
[MW] 23782.62
[pI] 6.18
[HOMOL] PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana 2e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YDR313c] 4e-06
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YOL013c]
0.001
[FUNCAT] 06.13 proteolysis [S. cerevisiae, YOL013c] 0.001
[PFAM] Zinc finger, C3HC4 type (RING finger)
[KW] Irregular

SEQ MGGKQSTAARSRGPFPGVSTDDSAVPPPGGAPHFHGYRTGGGAMGLRSRVSSVAGMGMD
PRD ccc
SEQ PSTAGGVFPGLYTPASRGTDGDSERAPGGGGSASDSTYAHNGYQETGGGHHRDGMLYLGS
PRD ccc
SEQ RASLADALPLHIAPRWFSHSGFKCPICSKSVASDEMEMHFIMCLSKPRLSYNDVLTGD
PRD hhhhhhhhhcecc
SEQ AGEVCICLEELQGDITARLPCLCIYHKSCIDSWFEVNRSCPEHPAD
PRD cceeeeecc

(No Prosite data available for DKFZphtes3_9e22.3)

Pfam for DKFZphtes3_9e22.3

HMM_NAME Zinc finger, C3HC4 type (RING finger)
HMM *CPICfCTFQlDyPWPfdePmMlPCgHsFCypCIrrW.....CPmC*
C IC L+++ D++ LPC+ ++ ++CI +W CP+
Query 184 CVIC-----LEELLQGDITARLPCLCIYHKSCIDSWFEVNRSCPEH 224

DKFZphtes3_9i20

group: testes derived

DKFZphtes3_9i20 encodes a novel 205 amino acid protein with similarity to human KIAA0336 gene.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="44.1 cR from top of Chr17 linkage group"

Insert length: 2509 bp

Poly A stretch at pos. 2499, polyadenylation signal at pos. 2481

```
1 CTCGCCGAGA TGACCTGGGC ACCTCTGCGT TGAATCGGCA AATACTGATC
51 AAGCCGCGATT TATTCTGCTC TCAGGAACCTC TAAGTCTAGC AGAGAAGATG
101 AGGCCGCTAGA AGTTCATCAA TGGCTTGGCT GGAGGACAAG CAAATTGAGG
151 ACATTGGCAA CGGAGTGATC AAAATGATAG ATCATGAGGC CTAATATGAA
201 TAAGGAAAGA AGAGAAGTGG CAGAGGCTGA GAACAGAAAG AGAGGGTGA
251 GGGGCTGTAA ATCTTGAAGA TTAGGGTATA ATATGAGTAT ATGGGTAAGA
301 ATTGGAAGAA TTGTGTAGGA GGCAGTAGTC AAAAAGTAGA AGCAGTTTGG
351 AAGAGTAGTT ACAAATATCA AGAGCCAGGT GGCTAAAAGG TGGAGCTATA
401 GGTCAATTGAA GCTCAAGAAA CTGAGTCTCT AGGGCATTGG TTAAGTCATC
451 TGTCTAGACT TCAAAGTTGT CTAGGATGAT AATCAGAAG ACTGATCTGT
501 GCCAAAGTCA CAGGTTTTTC ACGACTGAAA ACAACATAGC AAAATAAGCC
551 AAGATGTCTG TGGATCCAAT GACCTACGAG GCCCAGTTCT TTGGCTTCAC
601 CCCACAAACG TGCATGCTTC GGATCTACAT TGCATTTCAA GACTACCTAT
651 TTGAAGTGAT GCAGGCCGTT GAACAGGTTA TTCTGAAGAA GCTGGATGGC
701 ATCCCAGACT GTGACATTAG CCCAGTGCAG ATTCGCAAAAT GCACAGAGAA
751 GTTCTCTTGC TTCATGAAAG GACATTTTGA TAACCTTTTT AGCAAAATGG
801 AGCAACTGTT TTTGCAGCTG ATTTTACGTA TTCCCTCAA CATCTTGCTT
851 CCTGAAGATA AATGTAAGGA GACACCTTAT AGTGAGGAAG ATTTTCAGCA
901 TCTCCAGAAA GAAATTGAAC AGTTACAGGA GAAGTACAAG ACTGAATTAT
951 GTACTAAGCA GGCCCTTCTT GCAGAATTAG AAGAGCAAAA AATTGTTCAG
1001 GCCAAACTCA AACAGACGTT GACTTTCTTT GATGAGCTTC ATAATGTTGG
1051 CAGAGATCAT GGGACTAGTG ATTTTAGGGA GAGTTTAGTA TCCCTGGTTC
1101 AGAAGTCCAG AAAACTACAG AACATTAGAG ACAATGTGGA AAAGGAATCG
1151 AAACGACTGA AATATCTTA ATTGCTCAGT AGTCAAAAGG AGGAGCCTGT
1201 CAAAAGTAG AATCATAAGG ACTGTTCAAA CCATAAGGAC TGTCAAATC
1251 ATACCACTGA CTGTTCAAAC CAACCATACT TTTTATTAGA TTTGCTTTGT
1301 CAACCTCTTC TTGTATTCTG TGTTCCTCCT TTTTGTGGTC CACTTTGCTG
1351 AGGTATGAAG TGTACTACTT TGAAGTAGGC TGAAGCATCT GAGTCTTCTA
1401 ATAAGTGGA AGGGATCCAA CAAAGAAGCC ATGACCAGTT AAAGATATTT
1451 GCAGAGTTAC ACCTTGGTCA TAAGTCCTTT GTGACCTTGA TTATTTTGGC
1501 TTAATCTTTG GATGAGACCA GACAAGAAA GGATTAAACG GGTGGCTCCT
1551 TTAATATTAT TATTATTGTT TTTGAGACAA GGTCCCTTTC TGTACCCAG
1601 GTTAGAGTAG ATTTCACTGG CACAATCTTG GCTCACTGCA ACCTCTGTGT
1651 CCTGGGCTCA AGTGATCCTC CTGCCTCAGC CTCCCAAGTA GCTAGGACCA
1701 CAGGTGCGTG TCACCATGCT TGGCTAATTT TTTTGAGAA ACGAGGCCTC
1751 ACTATATTGT CCAGGCTGAG TGGCTCTTTT ATTAACCACT CATTACACTG
1801 CGGAACAGCC AACATAGAGT ACTTGCTCTC GTCCCTGTGA TTTTCTTTCA
1851 TGAGGGAGTC AATATGTAGT GGAAAGAAGC ATGTAGCAAA AAAGACAACC
1901 TTGATCTTTA ATAAAAAAGA AGTTGGTTTA TTTCCAAAT AAATCCCCTG
1951 ACAAAAAACC TGGTGATGTT AAGCAATTGA CTGTCTTAGA GTCCAGCAGA
2001 AGACCTTAGA CAAAAAAGC AGAACCCACT GGAGTAGAAA AGGAAGCATG
2051 TAGCATATAC TCAGTAGTGA AATTTAATTT TACTGACTGT TAGGTATCTA
2101 TGCCAAATTT TTTTCATACT TCAGTTGGTT TTGGAATCTG CCTTATACCT
2151 AATATTTATT TATTCACTAT CATAAGCATC AATATTTTAA TGCCCTCAGT
2201 GGGAAATTTG TGTTTAACT CAATGGAATC TAATATTTCT TTATGTCGTT
2251 AGTCCCTGTA AAATGTTAGG TCACCCAAGG AAAGGGGAGA AATAGCAATG
2301 GTTGTTCCTA AGGTATTGCT TGCCCTCCAT GTCTTCTTAA AGAGCAGAAC
2351 TTGGAGTTTC TCCTTTATGT AGAGAAGAAG TAACTTAGGG TGTATTGCA
2401 ATGAAATATT CATAGATATT GAAAGCTTGT GTTTACATGA AATATGTTTA
2451 TTATCAAGAA GTCCTTTTTT CAATTCTGTA CATTAAATAT ATGTGTTTTA
2501 AAAAAAAA
```

BLAST Results

Entry AC004148 from database EMBL:
Homo sapiens chromosome 17, clone HCIT524C5, complete sequence.
Score = 5245, P = 0.0e+00, identities = 1049/1049
3 exons

Entry HS556361 from database EMBL:
human STS TIGR-A003N29.
Score = 1005, P = 1.3e-39, identities = 201/201

Entry HSG043 from database EMBL:
human STS SHGC-36031.
Score = 955, P = 2.8e-37, identities = 205/215

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 554 bp to 1168 bp; peptide length: 205
Category: putative protein
Classification: no clue

```

1 MSVDPMTYEA QFFGFTPQTC MLRIYIAFQD YLFEVMQAVE QVILKKLDGI
51 PDCDISPVQI RKCTEKFLCF MKGHFDNLFS KMEQLFLQLI LRIPSNILLP
101 EDKCKETPYS EEDFQHLQKE IEQLQEKYKT ELCTKQALLA ELEEKIVQA
151 KLKQTLTFFD ELHNVGRDHG TSDFRESLVS LVQNSRKLQN IRDNVEKESK
201 RLKIS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9i20, frame 2

TREMBLNEW:HSAB2334_1 gene: "KIAA0336"; Human mRNA for KIAA0336 gene,
complete cds., N = 1, Score = 107, P = 0.0081

>TREMBLNEW:HSAB2334_1 gene: "KIAA0336"; Human mRNA for KIAA0336 gene,
complete cds.
Length = 1,583

HSPs:

Score = 107 (16.1 bits), Expect = 8.2e-03, P = 8.1e-03
Identities = 42/140 (30%), Positives = 76/140 (54%)

```

Query: 65 EKFLCFMKGHFDNLFSKMEQLFLQLILRIPSNILLPEDKCKETPYSEED----FQHLQKE 120
      EK CF+K H +NL +EQ +L R ILL +D ++P + D + L+++
Sbjct: 796 EKEKCFIkeh-ENLKPLLEQK--ELDRRAELILL-KDSLAKSPSVKNDPLSSVKEEEK 851

Query: 121 IEQLQE--KYKTELCTKQALLAELEEKIVQAKLKQTLTFFDELHNVGRDHGTSDFRESL 178
      IE L++ K K E K L+A ++ +K + + K+T T +EL ++ + S+
Sbjct: 852 IENLEKECKEKEEKINKIKLVA-VKAKKELDSSRKETQTVKEELESRLSEK--DQLSASM 908

Query: 179 VSLVQNSRKLQNIRDNVEKESKRLKI 204
      L+Q + +N+ EK+S++L +
Sbjct: 909 RDLIQGAESYKNLLLEYEKQSEQLDV 934

```

Pedant information for DKFZphtes3_9i20, frame 2

Report for DKFZphtes3_9i20.2

```

{LENGTH}      205
{MW}           24140.13
{pI}           5.51
{KW}           All_Alpha
{KW}           COILED_COIL      18.05 %

```

(No Pfam data available for DKFZphtes3_9i20.2)

DKFZphtes3_9k22

group: testes derived

DKFZphtes3_9k22 encodes a novel 304 amino acid protein with partial similarity to X. leavis katanin p80.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C-terminus of katanin p80

Sequenced by DKFZ

Locus: unknown

Insert length: 2676 bp

Poly A stretch at pos. 2665, no polyadenylation signal found

```
1 CTCTCTAGGC TGCCGGGCGC TGGTCGTCAG CGCCGAGGCT GGGCTGAGGC
51 GCCGCGGTAC CATGAGGCGC CGGTACTTAA GAGATTATGG CATCAGAAAC
101 CCACAATGTT AAAAAACGGA ACTTTTGTA TAAGATTGAG GATCATTTCA
151 TTGATCTTCC TAGAAAAAAG ATCTCTAATT TCACTAATAA GAACATGAAG
201 GAGGTTAAGA AATCTCCAAA ACAGTTGGCT GCTTACATAA ATAGAACAGT
251 TGGACAAACT GTGAAAGGCC CAGATAAACT TCGTAAAGTG ATCTATCGCA
301 GAAAGAAAGT TCATCATCCC TTTCCAAATC CTTGTTACAG AAAAAACAG
351 TCCCTGGAA GTGGGGGCTG TGACATGGCA AATAAGAAA ATGAAGTGGC
401 TTGTGACGCG CACCTGCCTG AAAAATTACA CCATGATAGT CGAACATATT
451 TGGTTAACTC CAGTGATTCT GGTCTTCAC AGACAGAAAG CCCATCATCA
501 AAATATAGTG GGTTTTTTTC TGAGGTTTCT CAGGACCATG AAACAATGGC
551 CCAAGTTTGT TCCAGCAGGA ATATGAGATT GAATGTAGCT TTAAGTTTCT
601 GGAGAAAGAG AAGTATAAGT GAAGTTGTAG CTTATTGTTT GAGGATAGAA
651 GATCTTGGCG TTGTGGTAGA TTGCCTTCCT GTGCTACCA ATTGTTTACA
701 GGAAGAAAAA CAATATATCT CACTTGGCTG CTGTGTTGAC TTGTTGCCCTC
751 TAGTAAAGTC ACTACTTAAA AGCAAAATTG AAGAATATGT TATAGTTGGT
801 TTAAGCTGGC TTCAAGCAGT CATTAAAGG TGGTGGTCAG AACTATCATC
851 CAAAACAGAA ATTATAAATG ATGGAATAT TCAAAATTTA AAACAACAAT
901 TAAGTGGATT ATGGGAACAG GAAAACCATC TTACTTTGGT TCCAGGATAT
951 ACTGGTAATA TAGCTAAGGA TGTAGATGCT TATTTATTAC AGTTACATTG
1001 AGAGATTICA TCTACTAAAG AGCATTGGT TTTTCAAAAC ATCCCTGAAC
1051 TGTATAATTT ACAAAAAATA AAGTCTCGTC TGAGAACTGT GAACTGTGGA
1101 AGAAATCAAA ACTATTTTTT CTTTAAATAA GCCACGTAAT GAAACCACTA
1151 ATGAAATCCC AGCAATCTGC TTCACATTGA AGTGGAAAAA TATCCAAAAG
1201 GAGCAGCTTC AATTTTCATTG AGGTGAAAGT GCACTATCAA GATTGTTTAC
1251 CTTTGTCTGA TTTGGGAGTT ATATGGTTAT TTGGTAACAT TAAGAACTAC
1301 TGGATTTTAA TGCAATCTCTG CATAAAAAATA TAATTTATAC TATGTGAAAA
1351 AATAAGACAG GACTTACCAC TAGGAACCAAC CAAGACCAAT CATCATTAAC
1401 TTTTAAAGA TTGTGTTTAA TTAAAAAAATA AAAACACTTA AATGTGTGCA
1451 GCTATTTTCT TATGTTGAAA AGACTGAAAG TTTAAACAT GAAAAAATC
1501 AATATTAAAC ATTTTTTGTG CACACTGAGA TACTGTGTAT GTAAATGCC
1551 TTAATTATTA ATAAGCCAAT GTGTTATGAT ACCAATATCT GTTTTAAAAA
1601 ACTAAACCA ACCATGCTTC TGGCATGATA AAATCATGGA ATTAATCAG
1651 GGGTTTACAT TCTGTAGAG TGTCTTGAA AACTCTCTG CACCATTTTT
1701 AAACTTGAG AATAGTTTAA GTATCTCTGA TATTTTTTGC CAGAATCATC
1751 ATGTCATGTA TGAATGTGTT ATCCCTATCT AAGGAAAAAG GTGAATATGT
1801 TTTTGTATGA ATGTTTAACT GGAATGTCC ATGGACTTGG CTAATTTATA
1851 TTTACTTTTT ATTGTACATA GATTCTAAT ATTTTTCATT CCTGTATCAT
1901 TTAACCTTCC TTCATTTGAG TAAATTCAT AAATATTCT ATTTTTTTGC
1951 TTTTTTAAAT TCTGATTTTA TATGAATTCT AATCTTTTT CACTACATAT
2001 GTTTTAAAGA GTTACATACA GTGATTTAGA ATGGTTTACA GTTAATGCTG
2051 ATCTTGATTT TTAATTTCCA AACTTTTGTG TCACTACCTC CTCTAATGGT
2101 TAGTATGATA TGCTAGCAGA CTGTATGAGG TCTTTTTTAA AAATACCACT
2151 TTTAGTGTCA GTGAACCAAA TTCTGGAATG TCTTAACAGC TCTAAATCTT
2201 ACTTGCTCTG AAAATGATTG GGGTTTAAATA CCACTGCTGG TGGTTCACAC
2251 ATCATCCCAT CCTTAATATG CCTGACAGGC ATCTGAGCAA AGGTTTTTAG
2301 TAATTGAATT TCTCTGCAGT AGTCCTTCAA GCCTTGAAT GTAAACCTTT
2351 AGCATTTATT CGTTAATGA CTAATGATAC GAATCTCAAG CAGATTTCTT
2401 GCTCTTAAAA GTTATGTTTC ACTGAGTTCT GGTTTTGTGT AGCTATATTT
2451 TATATAGCTA GATATTCCTC ACAGTGAACA TGAATTGTAA TAATTGGTTA
2501 TTTCCCTTAA TCTTTAGATT ATAATAATTT CAGATTATTG CACGCTCTGT
2551 ATTTGAGAGG TGAGTTATTT AAGAGGCCAG TTTTCAGGAC ATGGGAATTT
2601 GAATTGTAAA CCTGTATCT CTGTGAAACT TTTAACATGA TAAATATATA
2651 CCTTCTTTG TGCTTAAAAA AAAAAA
```

BLAST Results

Entry HS541354 from database EMBL:
human STS WI-11840.
Score = 1267, P = 7.1e-50, identities = 271/281

Medline entries

98227670:
Katanin, a microtubule-severing protein, is a novel AAA ATPase
that targets to the centrosome using a WD40-containing subunit.

Peptide information for frame 3

ORF from 87 bp to 998 bp; peptide length: 304
Category: similarity to known protein
Classification: unclassified

1 MASETHNVKK RNFCKIEDH FIDLPRKKIS NFTNKNMKEV KKSPKQLAAY
51 INRTVGQTVK SPDKLRKVIY RRRKVVHPPF NPCYRKKQSP GSGGCDMANK
101 ENELACAGHL PEKLHHSRT YLVNSSDSGS QTESPSSKY SGFFSEVSQD
151 HETMAQVLFS RNMRLNVALT FWRKRSISEL VAYLLRIEDL GVVVDCLPVL
201 TNCIQEEKQY ISLGCCVDLL PLVKSLLKSK FEEYVIVGLN WLQAVIKRWW
251 SELSSKTEII NDGNIQILKQ QLSGLWEQEN HLTLPVPGYTG NIAKDVDAYL
301 LQLH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9k22, frame 3

TREMBL:AF056021_1 product: "p80 katanin"; *Xenopus laevis* p80 katanin
mRNA, partial cds., N = 1, Score = 146, P = 1.2e-07

TREMBL:AF052432_1 product: "katanin p80 subunit"; *Homo sapiens* katanin
p80 subunit mRNA, complete cds., N = 1, Score = 150, P = 1.2e-07

TREMBL:AF052433_1 product: "katanin p80 subunit"; *Strongylocentrotus*
purpuratus katanin p80 subunit mRNA, complete cds., N = 2, Score = 146;
P = 4.2e-07

>TREMBL:AF052432_1 product: "katanin p80 subunit"; *Homo sapiens* katanin p80
subunit mRNA, complete cds.
Length = 655

HSPs:

Score = 150 (22.5 bits), Expect = 1.2e-07, P = 1.2e-07
Identities = 35/105 (33%), Positives = 55/105 (52%)

Query: 145 SEVSQDHETMAQVLFSRNMRLNVALTFWRKRSISELVAYLLRIEDLGVVVDCLPVLTNCL 204
S++ + H+TM VL SR+ L+ W I V + I DL VVVD L N +
Sbjct: 489 SQIRKGHDTMCVVLTSRHKNLDTVRVWMTMGDIKTSVDSAVAINDSLVSVDLL----NIV 544
Query: 205 QEEKQYISLGCCVDLLPLVKSLLKSKFEEYVIVGLNWLQAVIKRW 249
++ L C +LP ++ LL+SK+E YV G L+ +++R+
Sbjct: 545 NQRASLWKLDLCTTVLPQIEKLLQSKYESYVQTGCTSLKLILQRF 589

Pedant information for DKFZphtes3_9k22, frame 3

Report for DKFZphtes3_9k22.3

{LENGTH} 304
{MW} 34767.24
{pI} 9.18
{KW} All_Alpha

```

SEQ      MASETHNVKKRNFNCNKIEDHFDLPRKKISNFTNKNMKVEVKKSPKQLAAYINRTVGQTVK
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccc

SEQ      SPDKLRKVIYRRKKVHHFFPNPCYRKQSPGSGGCDMANKENELACAGHLPEKLHHSRI
SEG      .....
PRD      cchhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhcccccccccccccc

SEQ      YLVNSSDSGSSQTESPSSKYSGFFSEVSQDHETMAQVLFSRNMRLNVALTFWRKRSISEL
SEG      .....
PRD      eeeeeccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      VAYLLRIEDLGVVVDCLPVLTNCLQEERQYISLGCCVDLLPLVKSLLKSFEYVIVGLN
SEG      .....
PRD      hhhhhhhhhccccccccccccchhhhhhhhhccccccccceehhhhhhhhhhhheeeeeehh

SEQ      WLQAVIKRWWSELSSKTEIINDGNIQILKQQLSGLWEQENHLTLVPGYTGNIADVDAYL
SEG      .....
PRD      hhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhccccccccchhhhhhhh

SEQ      LQLH
SEG      ...
PRD      hccc

```

(No Pfam data available for DKFZphtes3_9k22.3)

Localization of expressed proteins

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp434B0435	AL117496	6248	48	5366	1773	transport and traffic	similar to: kinesin like proteins	512.1 cR from top of Chr10 linkage group	"secr pathway"	None
DKFZp434N0535	AL117518	4055	126	4025	1300	differentiation & development	similar to: Drosophila chromatin protein		"no predict"	None
DKFZp564A0122	AL110209	2722	65	1300	412	signaling & communication	similar to: acyltransferase	16	"mitochondria"	Mitochondria
DKFZp564A022	AL136620	1376	132	632	167	unknown	unknown	4	"no predict"	Endoplasmic Reticulum
DKFZp564A032	AL50267	2214	76	1953	626	differentiation & development	similar to: MG21 contains three conserved protein motifs present in GTP-binding proteins, but these are not conserved in 2_2a3.1	238.7 cR from top of Chr20 linkage group	"no predict"	Nucleus
DKFZp564A0723	AL80116	2524	42	2177	712	cell cycle	similar to: origin recognition complex	6q14.3-16.1	"nucleus"	Cytosol + Nucleus
DKFZp564A202	AL80056	707	33	620	196	metabolism	similar to: protein involved in energy metabolism	10	"mitochondria"	Endoplasmic Reticulum
DKFZp564B0482	AL110243	2092	317	1579	421	signaling & communication	Unknown, contains 2 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins		"no predict"	Cytosol + Nucleus
DKFZp564B1023	AL136611	2905	157	1896	580	nucleic acid management	similar to: RNA helicase		"nucleus / nuclear envelope"	Nucleus

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKF2p564B1162	AL136646	4593	661	2625	655	signaling & communication	unknown	4	"no predict"	Cytoskeleton
DKF2p564B122	AL049972	1842	70	1536	488	unknown	unknown		"no predict"	Cytosol
DKF2p564B1471	AL136667	1484	78	323	82	membrane protein	unknown		"no predict"	Endoplasmic Reticulum
DKF2p564B162	AL136621	1914	246	1631	462	nucleic acid management	similar to: Zinc finger protein	13q12	"no predict"	other/unknc
DKF2p564B163	AL50268	1208	191	577	129	transport and traffic	similar to: GTP binding protein		"secr pathway"	Cytosol + Nucleus
DKF2p564B212	AL136623	1915	218	1348	377	protein management	similar to: protein involved in posttranslational modification	22q12.1	"secr pathway"	Endoplasmic Reticulum
DKF2p564B2123	AL136612	3300	121	699	193	signaling & communication	similar to: Neurocalcin is a Ca(2+)-binding protein with 3 EF-hands. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.	574.6 CR from top of Chr8 linkage group	"no predict"	Cytosol + Nucleus
DKF2p564B246	AL136664	2054	73	1074	334	metabolism	similar to: dTDP-6-deoxy-L-mannose-dehydrogenase	5	"nucleus or cytosol"	Nucleus
DKF2p564C0362	AL80076	1731	60	1142	361	nucleic acid management	similar to: ssDNA binding protein		"no predict"	other/unknown
DKF2p564C0469	AL050298	899	86	898	270	unknown	unknown		"nucleus"	Cytosol + Nucleus
DKF2p564C1362	AL136647	837	137	673	179	metabolism	similar to: molecular clock protein	16p12.3-p13.11	"mitochondria"	Other/unknown

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564C1616	AL136597	3928	240	1997	586	structure & motility	shares the features of mayven and kelch and therefore should be involved in the organisation of cytoskeleton binding to membrane proteins		"cytoskeleton / plasma membrane"	Nucleus
DKFZp564C162	AL136627	2305	155	625	157	membrane protein	unknown	86.2 cR from top of Chr1 linkage group	"no predict"	Endoplasmic Reticulum
DKFZp564C1664	AL136656	1866	180	1040	287	unknown	unknown	745_A_2; 756_F_2; 842_C_2	"no predict"	Cytosol
DKFZp564C182	AL136628	2835	272	1177	302	unknown	unknown		"no predict"	Golgi
DKFZp564C183	AL136639	1709	105	1448	448	nucleic acid management	similar to: DEAD-box helicase	87.50 cR from top of Chr16 linkage group	"nucleus / nuclear envelope"	Nuclear envelope
DKFZp564C196	AL050020	2266	366	966	200	signaling & communication	similar to: neuronal calcium sensor		"no predict"	Nucleus
DKFZp564D116	AL050022	2535	29	1849	607	signaling & communication	similar to: GTP-binding protein		"no predict"	Cytosol
DKFZp564D202	AL136631	1787	18	944	309	unknown	unknown		"no predict"	Cytosol
DKFZp564E0123	AL136613	2005	104	1000	299	unknown	unknown	16q13	"no predict"	Cytosol + Nucleus
DKFZp564E0482	AL136697	2923	163	1581	473	signaling & communication	similar to: calmodulin-related protein	200.5 cR from top of Chr3 linkage group	"no predict"	Cytosol + Nucleus

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564E1782	AL136696	1618	40	972	311	membrane protein	unknown	171.7 cR from top of Chr14 linkage group	"no predict"	Endoplasmic Reticulum
DKFZp564E2182	AL50261	2367	193	804	204	Cell Cycle	similar to: protein involved in cell cycle, DNA repair, maintenance of minichromosomes	6q22.1-22.33	"nucleus"	Nucleus
DKFZp564F0223	AL136614	1016	68	613	182	unknown	unknown	12q24	"secret pathway / endosomes"	other/unknown
DKFZp564F052	AL049989	1649	34	1303	423	signaling & communication	similar to: sorting nexin 7		"membranes"	Cytosol
DKFZp564F0522	AL049943	2078	283	943	220	unknown	unknown	2	"no predict"	Nucleus
DKFZp564F1862	AL80081	1987	250	918	223	differentiation & development	similar to: DnaJ proteins, but lacks CRR domain of these proteins.	7q31	"no predict"	Endoplasmic Reticulum
DKFZp564F2116	AL136598	1512	115	738	208	membrane protein	unknown	15q25	"nucleus"	other/unknown
DKFZp564F2122	AL136604	1910	156	1856	567	unknown	unknown	311.4 cR from top of Chr14 linkage group	"no predict"	Cytoskeleton (microtubules)
DKFZp564F2162	AL136648	1549	95	730	212	unknown	unknown	209.8 cR from top of Chr20 linkage group	"peroxisomes"	Peroxisomes
DKFZp564G0222	AL80115	1165	157	933	259	nucleic acid management	unknown		"no predict"	Endoplasmic Reticulum

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKF2p564G083	AL136641	1027	37	570	178	protein management	similar to: yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1.	20	"no predict"	Cytosol + Nucleus
DKF2p564G182	AL136632	2444	539	1225	229	unknown	unknown	6p22.1-22	"no predict"	Cytosol + Nucleus
DKF2p564H012	AL136633	957	93	632	180	unknown	unknown		"no predict"	Mitochondri
DKF2p564H1122	AL136605	1734	159	1133	325	membrane protein	unknown	11q14	"no predict"	Nucleus
DKF2p564H1322	AL136606	2292	270	1829	520	membrane protein	unknown	19q13.2 from BCKDHA-D19S217	"no predict"	Cytosol + Nucleus
DKF2p564H1562	AL136649	2014	75	971	299	structure & motility	similar to: Cell cell interaction protein	1	"plasma membrane"	plasma membrane + cell contact sites
DKF2p564I0123	AL136615	1467	126	1064	313	signaling & communication	similar to: protein activator of the interferon-induced protein kinase		"cytosol or nucleus"	Cytosol
DKF2p564I0422	AL136607	4748	511	1194	228	signaling & communication	unknown		"no predict"	Golgi + Plasma membrane
DKF2p564I1216	AL136600	1548	81	635	185	membrane protein	unknown	873.3-875.1 CR from top of Chr1 linkage group	"no predict"	Endoplasmic Reticulum
DKF2p564I1782	AL136699	1741	168	410	81	signaling & communication	similar to: phospholemman protein,	11q23	"secr pathway"	Golgi + plasma

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
							a membrane substrate for the cAMP-dependent protein kinase; seems to serve as chloride channels or as chloride-channel regulators. Transmembrane protein			membrane
DKF2p564I206	AL136665	1122	34	921	296	unknown	unknown	377.5 cR from top of Chr8 linkage group	"mitochondria"	Mitochondria
DKF2p564I2423	AL136616	1713	58	882	275	metabolism	similar to: protein involved in amino acid metabolism	8p11.2	"cytosol"	Cytosol + Nucleus
DKF2p564I2482	AL136700	1860	10	1650	547	nucleic acid management	similar to: Dead-box helicase	175.5 cR from topFT of Chr7 linkage group	"nucleus / nuclear envelope"	Nucleus
DKF2p564J1022	AL110301	1409	5	1021	290	nucleic acid management	Unknown, contains a Leucine zipper	12	"cytosol or nucleus"	Cytosol + Nucleus
DKF2p564J1516	AL136601	2868	352	1839	496	structure & motility	similar to: RNA binding, Tubulin binding	20, 12.10 cR from GCT10F11	"cytosol"	Cytosol
DKF2p564J1864	AL136660	690	109	648	180	transport and traffic	similar to: canin and chicken microsomal signal peptidase 23 kd subunit.		"endoplasmic reticulum"	Endoplasmic Reticulum
DKF2p564J2222	AL136608	1858	154	1440	429	structure & motility	similar to: actin-related protein		"plasma membrane / cytoskeleton"	plasma membrane
DKF2p564K0322	AL136609	2775	779	2392	538	unknown	unknown		"no"	Plasma

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
									predict"	membrane
DKF2p564K0822	AL136610	2789	10	525	172	unknown	unknown	7	"no predict"	Golgi
DKF2p564K1216	AL49933	1938	357	1418	354	signaling & communication	similar to: GTP-binding regulatory protein	7	"membranes"	Golgi + Plasma membrane
DKF2p564K192	AL136637	1931	107	1015	303	unknown	unknown	6p22.1-22.3	"no predict"	Other/unknown
DKF2p564K1964	AL117619	1560	207	884	226	unknown	unknown	17	"no predict"	Endoplasmic Reticulum
DKF2p564K2216	AL136602	2088	832	1155	108	unknown	unknown		"no predict"	Mitochondria
DKF2p564L023	AL136643	2978	279	2045	589	protein management	Unknown, Pfam prediction: ubiquitin family	9	"cytosol"	Cytosol + Nucleus
DKF2p564L1216	AL136603	2042	73	873	267	membrane protein	unknown		"secre pathway"	Golgi + plasma membrane
DKF2p564L2423	AL136617	2416	29	1072	348	transport and traffic	Unknown, a lectin character is predicted	2	"endoplasmic reticulum"	Endoplasmic Reticulum
DKF2p564M082	AL80071	902	227	589	121	unknown	Unknown, contains osteopontin motive		"no predict"	Cytosol + Nucleus
DKF2p564M112	AL80070	2686	14	595	194	signaling & communication	unknown	956.7 cR from top of Chr2 linkage group	"no predict"	Golgi
DKF2p564M173	AL136644	636	26	400	125	unknown	similar to: janus proteins		"no predict"	Cytosol + Nucleus
DKF2p564M1863	AL117602	1192	125	1027	301	signaling &	similar to: phospho- like protein, G-protein	9	"cytosol"	Cytosol

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
						communication	modulator			
DKFZp564M1982	AL390217	2707	302	1160	286	unknown	unknown		"no predict"	Cytosol
DKFZp564M2423	AL80119	2201	86	1246	387	unknown	unknown	72.60 cR from top of Chr3 linkage group	"cytosol"	Cytosol
DKFZp564N0582	AL50264	1646	75	506	144	cell cycle	similar to: DRR1 gene	3p21.1	"cytoskeleton / plasma membrane"	Cytoskeleton (focal adhesion sites) + nucleus
DKFZp564N1623	AL136618	2936	172	1047	292	signaling & communication	Unknown, contains a WW domain which binds proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes		"no predict"	Cytosol + Nucleus
DKFZp564O043	AL050390	2515	186	1509	441	structure & motility	similar to: ankyrin	7	"no predict"	Cytosol + Nucleus
DKFZp564O0523	AL136619	1736	24	1103	360	unknown	unknown	7q21-q22	"no predict"	Nucleus
DKFZp564O123	AL80122	1985	234	872	213	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp564O1762	AL136652	1260	56	901	282	signaling & communication	similar to: low-density lipoprotein (LDL) receptors are the major		"secretory pathway"	Golgi

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
							cholesterol-carrying lipoproteins of plasma. The novel protein contains an additional leucine zipper suitable for protein-protein interaction.			
DKFZp56401923	AL050295	2091	237	2090	617	metabolism	similar to: dTDP-6-deoxy-L-mannose-dehydrogenase		"secr pathway"	Cytosol
DKFZp56402423	AL390214	3564	656	1072	139	unknown	Unknown, contains CAAX box (prenyl group binding site); found in :Ras proteins, and Ras-like proteins such as Rho, Rab, Rac, Ral, and Rap; nuclear lamins A and B; Some G protein alpha subunits, G protein gamma subunits; some dnaJ-like proteins		"no predict"	Cytosol + Nucleus
DKFZp5640243	AL050015	1074	23	834	270	unknown	unknown	3	"no predict"	Endoplasmic Reticulum
DKFZp56611024	AL050037	1783	5	970	322	unknown	similar to: hypothetical protein Rv0712 - Mycobacterium tuberculosis		"no predict"	Cytosol
DKFZp566J2046	AL136720	1706	16	678	221	metabolism	similar to: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.	16	"no predict"	Mitochondria
DKFZp566K144	AL136727	3084	456	1079	208	transport and traffic	similar to: Rab protein		"secr pathway"	Golgi
DKFZp586D0919	AL050100	2777	48	494	148	unknown	unknown	12	"no"	Golgi

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
									predict"	
DKF2p586E1124	AL136942	2005	184	861	226	transport and traffic	similar to: golgi transmembrane spanning transporter	8	"Golgi"	Golgi + plasma membrane
DKF2p586E1323	AL136936	1854	367	954	196	protein management	similar to: heat shock protein	578.9 cR from top of Chr12 linkage group	"cytosol or nucleus"	Cytosol + Nucleus
DKF2p586E1519	AL050101	2140	82	1680	559	unknown	similar to: A.thaliana A_IG002N01		"no predict"	Cytosol
DKF2p586F1918	AL050091	3489	184	594	137	unknown	unknown		"no predict"	Cytosol + Nucleus
DKF2p586F1919	AL136915	2024	134	745	204	membrane protein	unknown	14.8 cR from top of Chr20 linkage group	"secr pathway"	Golgi + plasma membrane
DKF2p586H2219	AL50282	1971	199	1623	475	unknown	unknown	22q11.2-qter	"no predict"	Cytosol
DKF2p586I0418	AL136912	1568	163	822	220	unknown	unknown	7q31	"no predict"	Cytosol + Nucleus
DKF2p586I1520	AL050149	2439	11	1711	566	transport and traffic	similar to: nuclear RanGTP binding protein		"nucleus"	Nucleus
DKF2p586J1023	AL136938	1048	72	749	226	protein management	similar to: glutathione S-transferase / posttranslational modification		"no predict"	Cytosol + Nucleus
DKF2p586J1119	AL136919	2343	28	2151	708	signaling & communication	unknown		"membranes"	Endoplasmic Reticulum
DKF2p586J1923	AL050220	745	49	588	179	differentiation &	similar to: serine protease	19	"secr pathway"	Endoplasmic Reticulum

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
						development				
DKFZp586K0919	AL50283	1782	204	1316	371	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp586L0118	AL136913	1076	45	596	184	protein management	similar to: mitochondrial Ribosomal S40 protein		"nucleus"	Nucleus
DKFZp586M2420	AL136927	1986	23	1855	611	transport and traffic	similar to: mannosyltransferase	11	"secr pathway"	Endoplasmic Reticulum
DKFZp727E151	AL390215	1957	340	1701	454	membrane protein	similar to: transporter proteins (contains 9 transmembrane domains)		"no predict"	Endoplasmic Reticulum
DKFZp727M111	AL117479	2275	79	1899	633	unknown	unknown		"no predict"	Cytosol
DKFZp727M231	AL117480	2428	56	1681	542	unknown	unknown		"no predict"	Cytosol
DKFZp761G05121	AL118986	4592	107	3613	1169	protein management	similar to: SH3 BINDING PROTEIN		"cytosol"	Cytosol
DKFZp761G18121	AL136548	4117	107	1438	444	signaling & communication	similar to: ALLOGRAFT INFLAMMATORY FACTOR		"nucleus"	Nucleus
DKFZp761I12121	AL136549	4130	139	3894	1252	cell cycle	similar to: p53 inducible protein	5q34	"no predict"	Cytosol
DKFZp761M02121	AL136551	3328	178	2163	662	cell cycle	similar to: p53 regulated PA26-T2 nuclear protein		"nucleus"	Cytosol
DKFZp761O15121	AL136552	4293	112	2421	770	signaling & communication	similar to: semaphorin W	328.8 cr from top of Chr2 linkage group	"secr pathway"	Endoplasmic Reticulum

Table of cDNA clones and related data

Group: cell cycle

Clone ID	Homology	Function	Group
hbr2_16g18	Similarity to KIAA0797 and yeast Smt4p	Novel protein with similarities to S. pombe SPAC17A5.07c and the S. cerevisiae Smt4p suppressor of Mif2 gene.; involved in centromere organisation	Cell cycle
hbr2_2k14	Strong similarity to human N33 tumour suppressor gene	New tumour suppressor gene	Cell cycle
htes3_35b4	Human M-phase phosphoprotein-1	The novel protein is C-terminal identical to human M-phase phosphoprotein-1, which is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.	Cell cycle
htes3_35p22	Strong similarity to oncogene 1 (trc-2 locus)	Oncogene	Cell cycle
htes3_7j3	Related to the C-TAK1 Cdc25C associated protein kinase	Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five	Cell cycle
htes3_7p10	Strong similarity to XPMC2 protein	XPMC2 of xanopus rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function.	Cell cycle
hut1_20m11	Similarity to suppressor protein sds22	Suppressor regulator of protein phosphatase-1	Cell cycle

Group cell structure and motility

Gene ID	Homology	Function	Group
hfr2_16c16	Similarity to Drosophila kelch	Shares the features of mayven and kelch and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins	Structure and motility
hfr2_2b5	Similarity to collagen proteins	New collagen alpha chain	Structure and motility
htes3_1515	Strong similarity to "radial spokehead" proteins	Part of sperm motor	Structure and motility
htes3_1817	Similarity to ankyrins	Putative ankyrin	Structure and motility
htes3_1k11	Strong similarity to mouse ENC-1	Nuclear matrix protein	Structure and motility
htes3_72k15	Strong similarity to Rattus norvegicus actin-filament binding protein Frabin.	FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin binding protein. Modulation of cell structure and motility as well as modulation of the JNK/SAPK pathway.	Structure and motility
htes3_7b22	Similarity to paramyosins	Protein involved in motility	Structure and motility
hut1_19g22	Strong similarity to tuftelin/enamelin	New connective tissue protein	Structure and motility
hut1_24j6	Strong similarity Rattus norvegicus cell adhesion regulator (CAR1) mRNA	Cell adhesion regulator (signal transduction molecule influencing cell adhesion to collagen)	Structure and motility

Group Differentiation/Development

CloneID	Homology	Function	Group
hibr2_2d15	Mus musculus testis-specific Y-encoded-like protein (tspy11).	TSPY is believed to function in early spermatogenesis and is a candidate for GSY, the putative gonadoblastoma-inducing gene on the Y-chromosome	Differentiation/Development
htes3_35e21	Similarity to interleukin-7 precursor	New interleukin	Differentiation/Development
hut1_2h3	Strong similarity to mouse E25 and gallus E3-16	Homolog is marker for chondro-osteogenic differentiation	Differentiation/Development

Group kidney derived

CloneID	Homology	Function	Group
hfk2_1j9	Strong similarity to XLCL2 protein, African clawed frog	No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_24e23	Unknown	No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_46a6	Unknown	No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_46b10	Similarity to C.elegans F25B5.3	No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_46d13	Weak similarity to KE03 protein	Contains a RGD site; No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_4b6	Similarity to Homo sapiens clone 25003 partial CDS.	No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_4c8	Similarity to KIAA0549 and HAP1 (Huntingtin-associated protein-1)	No Informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived

Group mammary carcinoma derived

Clone ID pyp	Homology	Function	Group
hmcfl_lc23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Mammary Carcinoma derived
hmcfl_lgl3	Similarity to KIAA0766; very weak similarity to transposases	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Mammary Carcinoma derived

Group Nucleic acid management

CloneID DEF	Accession Accession	Gene Gene	Function Function	Group Group
hfr2_23b10	HEL117	Similarity to rat RNA helicase	RNA helicase	Nucleic Acid Management
hfr2_3c18		Strong similarity to RNA helicase and RNA-dependent ATPase from the DEAD box family	DEAD-box	Nucleic Acid Management
hfr2_64a15		Similarity to inorganic pyrophosphatases (unspliced)	Inorganic pyrophosphatase	Nucleic Acid Management
hfr2_6017		Strong similar to RNA helicases	RNA helicases	Nucleic Acid Management
hfr2_72b18		Similarity to DNA damage induced genes	Similar to dinp of <i>E. coli</i> , YqjH of <i>B. subtilis</i> , dinp of <i>M. tuberculosis</i> and T19K24.15 of <i>A. thaliana</i> . The dinB/P pathway is a second SOS-pathway in <i>E. coli</i>	Nucleic Acid Management
hfr2_72112		Similarity to YDR126W	DNA binding protein	Nucleic acid management
hfr2_82124		Strong similarity to DEAD-box subfamily ATP-dependent helicase	Dead-box helicase	Nucleic Acid Management
htes3_14h21		Strong similarity to RNA helicases	RNA helicase	Nucleic Acid Management
htes3_15j3		Similarity to YGR276C, a ribonuclease H of <i>S. cerevisiae</i> .	Rnase H	Nucleic Acid Management
htes3_20m18		Similarity to the <i>S. cerevisiae</i> mitochondrial carrier protein RIM2.	The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer.	Nucleic Acid Management
htes3_22g2		XIA0829 is shorter, nearly identical to rat TIP120	Involved in TATA box binding complex	Nucleic Acid Management
htes3_2m18		Nearly identical to mouse Dhml	Multifunctional nuclease/exoribonuclease	Nucleic acid management
htes3_7p9		Similarity to nuclear domain 10 protein NDP52	Transcription control	Nucleic Acid Management
htes3_8m10		Strong similarity to polyadenylate-binding proteins.	The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.	Nucleic Acid Management
hute1_18l1		Strong similarity to <i>S. cerevisiae</i> YHR148W	Mitochondrial Ribosomal S40 protein	Nucleic Acid Management

Group testis associated

Cloning Data	Homology	Function	Group
htes3_14g5	Strong similarity to cell growth regulating nucleolar protein LYAR, of mouse	Contains a ATP/GTP-binding site motif A (P-loop), but not the zinc finger motif and nuclear localization signals of lyar.	Testes associated
htes3_14p14	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_14p7	Weak similarity to kinesin associated protein KAP3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15a13	Similarity to S.cerevisiae Hop1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15g14	Similarity to YOR243c	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15h1	Weak similarity to Hsp70/Hsp90 organizing protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15j18	Unknown	Unknown; no predictive prosite pfam or SCOP motive	Testes associated
htes3_17f10	T23R7.2B PROTEIN	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_18f3	Similarity to TNF-inducible protein CG12-1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_19f19	Weak similarity to S. cerevisiae protein YPL046w.	The protein contains a RGD cell attachment site.	Testes associated
htes3_19j17	Partial similarity to C.elegans Y40B1A.2 protein.	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_20c21	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_21n23	Strong similarity to rat 7a comp protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_22c23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_22n13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_27o14	Similarity to C.elegans C55A6.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_28d14	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2a11	Similarity to mucin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2d15	Similarity to C.elegans F25H2.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2f14	Weak similarity to omega protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2g7	Similarity to neurofilament proteins	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2h15	Similarity to S.pombe cdc23	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2l19	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated

CloneID	Homology	Function	Group
htes3_2m20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_2n9	Very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_30f4	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_35g6	Strong similarity to R27216_1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_35n24	Unknown	Contains an IG_MHC pattern	Testes
htes3_35p17	Similarity to S.cerevisiae VAC8 and beta-Catenin, but contains no amadillo motifs	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4b4	Rattus norvegicus late gestation lung protein 1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4f17	Similarity to KIAA0333	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4o19	Methyl-CpG binding protein; does not contain such a motive.	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50j4	Similarity to mucin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50n23	Unknown, prolin rich protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50n6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_6b21	Similarity to KIAA0256	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_6d16	WUGSC:H DJ1185I07.2, differences to genmodel	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_72k11	Similarity to S.pombe hypothetical repeat-containing protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_7d17	Similarity to KIAA0454	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_7j8	WUGSC:H DJ1159O04.1 similarity to YBL104p	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8g11	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8g5	KIAA087, alternative spliced	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8p7	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9e22	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9i20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9k22	Similarity to C-terminus of katanin p80	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated

Group transmembrane proteins

Clonaid	Homology	Function	Group
hfr2_16112	Similarity to Fugu rubripes PUT2	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Transmembran e protein
hfr2_16112	Similarity to gallus putative transmembrane protein E3-16	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_22h13	Similarity to Drosophila melanogaster EG:39E1.3.	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_2b17	Similarity to Drosophila hypothetical 30K protein	3 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_2d17	Unknown	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_64k24	Similarity to several proteins	5 transmembrane regions. No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Transmembran e protein
hfr2_82c20	Similarity to C.elegans D1007.5	7 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_82e17	Similarity to C.elegans "R01B10.5"	6 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_82g14	Unknown proline rich protein	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_24a15	Similarity to C. elegans R07G3.8	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_3113	Similarity to A.thaliana YUP8H12.2	3 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hfr2_4m11	Weak similarity to YMR034c	4 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hmcfl_lal1	Similarity to YDR255c and SPBC29A3.03c	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hmcfl_le15	Similarity to D-XYLOSE TRANSPORTER	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
htea3_15c6	Unknown	Transporter; 9 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
htea3_2ol3	Partial similarity to the IL-17 receptor.	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
htea3_27k4	Strong similarity to C.elegans K07H8.2/2K185.2	1 transmembrane domain Contains a leucine zipper 10 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive;	Transmembran e protein
htea3_2h1	Similarity to C.elegans C13P10.5	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
htea3_35k24	Unknown	5 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hutel_19f19	Similarity to mouse P24 protein	2 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein
hutel_24c19	Unknown	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembran e protein

Group Brain derived

GeneID	Homology	Function	Group
hbr2_16f21	Strong similarity to zinc finger protein 216 has no zn finger, is only similar	PROSITE: Contains no Zinc finger; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_16k22	Weak similarity to thioredoxin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_22f21	Weak similarity to C.elegans C18C4.5	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_22i4	Similarity to Human P52rIPK N-terminus	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_22k3	Weak homology with : EXTENSIN (PROLINE-RICH GLYCOPROTEIN)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_22k8	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_23f2	Similarity to Vp29p; saccharomyces cerevisiae (baker's yeast) pep11 protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_23o24	Similarity to CAAK-box protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_23o5	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_242	Similarity to 52K autoantigen Ro/SS-A - human	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i8	Weak similarity to cyclin-dependent kinase p130-PITSLRE	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i9	Similarity to Synechocystis sp. (PCC 6803)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i8	J30M3.2 extension of genmodel	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i1	Similarity to C.elegans D2007.4 protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i0	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_24i9	Similarity to KIAA0378	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_3f16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hbr2_3i2	Weak similarity to ubiquitin-like protein DSK2 yeast	Pfam: ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive	Brain derived
hbr2_62n10	Similarity to reticulocyte-binding protein	Contains a Leucine zipper; No informative BLAST results; No predictive pfam or SCOP motive	Brain derived
hbr2_64a11	Similarity to Drosophila irregular chiasm C-roughest precursor (frame shift)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived

CloneID Accession	Homology	Function	Group
hfbr2_64c16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64c4	Similarity to A. thaliana T08113.5	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64h6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64i20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64o16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_6a17	Weak similarity to finger protein zfoc1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_6i20	Similarity to ribosomal protein L15 precursor, mitochondrial	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_71o20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_72d13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_72m16	Similarity to C.elegans H14A12.3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_72n12	Strong similarity to rat Ganglioside expression factor (GEP-2) but even higher identity with C.elegans putative protein identities = 91/116 (78t)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_76d13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_78n23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_7a24	DKF2phfbr2_7a24.1 similarity to C- terminus of TGF-beta-activated kinase	Only c-terminus homolog; contains no kinase domain; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_7e22	Similarity to cytochrome b561	No heme domain but a c may helix loop helix signature No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_7j4	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_82m16	Very weak similarity to A.thaliana P28A23.140	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived

Group Intracellular Transport and Trafficking

Accession	Homology	Function	Group
hbr2_23124	Strong similarity to human GP36b glycoprotein and canine VIP 36	A lectin character is predicted. Due to the intracellular localisation of the homologue proteins, it should be involved in cell trafficking	Transport and traffic
hbr2_2117	Strong similarity to rab1	GTP binding	Transport and traffic
hbr2_41m15	Strong similarity to ras-related GTP-binding protein Rab17	GTP-binding, signal transduction	Transport and traffic
hbr2_62f10	Strong similarity to zinc transporter proteins	Zinc transporter protein	Transport and traffic
hbr2_62119	2 nearly identical to dog GTP-binding protein rab10	GTP binding	Transport and traffic
hbr2_64j18	Strong similarity to dog signal peptidase (EC 3.4.99.-)	Identical to canin and chicken microsomal signal peptidase 23 kd subunit.	Transport and traffic
hkd2_24n20	Strong similarity to eps8 binding protein e3B1	Contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e3B1) and spectrins. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton	Transport and traffic
hkd2_24p5	Human ankyrin G (ANK-3) new splice variant	New ankyrin protein	Transport and traffic
hkd2_4k14	Strong similarity to Rab6	New Rab protein	Transport and traffic
htes3_1g13	Similarity to 256 kb golgin, strong similarity to rat "cp151"	New golgin protein	Transport and traffic
htes3_17n18	TonB-dependent receptor protein signature 1	Involved in receptor-mediated uptake	Transport and traffic
htes3_21116	Identical to rat ribosome attached membrane protein 4	Responsible for transport of proteins into ER	Transport and traffic
htes3_23111	Nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).	Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors	Transport and traffic
htes3_26g22	Similarity to kinesins.	The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport.	Transport and traffic
htes3_4h6	Strong similarity to Kinesin light chain	New kinesin light chain	Transport and traffic
htes3_72p16	Strong similarity to mouse MEM3 and yeast VPS35	New vacuolar protein sorting-associated protein	Transport and traffic
hutel_19h17	Strong similarity to C.elegans ZK1086.1	Steroid turnover in cells	Transport and traffic
hutel_20h13	Strong similarity to alpha-adaptins	New adaptin chain (clathrin assembly protein complex 2 alpha-a large chain)	Transport and traffic
hutel_24e11	Similarity to golgi 4-transmembrane spanning transporter mtp	New golgi transmembrane spanning transporter	Transport and traffic

Group signal transduction

CloneID	Homology	Function	Group
hfb2_23b21	Nearly identical to bovine neurocalcin	Neurocalcin is a Ca(2+)-binding protein with 3 EF-hands. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.	Signal transduction
hfb2_23n16	Similarity to putative phosphatidylinositol-4-phosphate 5-kinase	Contains a WW domain which binds proteins with particular proline- motifs, [AP]-P- (AP)-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes	Signal transduction
hfb2_2c17	(similarity to YMR131c and retinoblastoma-binding protein RBAP46)	The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins.	Signal transduction
hfb2_2b11	Putative GTPase-activating protein, related to human chimaerins	The new protein is expected to activate p21rac-related small GTPases	Signal transduction
hfb2_78c24	Strong similarity to guanylate-binding proteins (GBPs)	Modulating/blocking the response of cells to interferons.	Signal transduction
hfb2_82e4	Strong similarity to rat calmodulin-binding protein	Involved in calmodulin-related pathway	Signal transduction
hfb2_82f17	Similarity to plasma membrane substrate for cAMP-dependent protein kinase	Transmembrane protein with strong similarity to the phospholipase protein, a membrane substrate for the cAMP-dependent protein kinase; seems to serve as chloride channels or as chloride-channel regulators. Transmembrane Protein	Signal Transduction
hfb2_82m6	Strong similarity to mouse "sphingosine kinase	Sphingosine kinase	Signal transduction
hfk2_46m4	Nearly identical to mouse GTP-binding protein	GTP-binding protein	Signal transduction
htes3_15k11	KIAA0781, 5' extension	Heart development/signal transduction	Signal transduction
htes3_1c1	Similarity to GTPase-activating proteins	GTPase-activating proteins	Signal transduction
htes3_1n3	Similarity to Tup1p	Beta-transducin subunit of G-proteins	Signal transduction
htes3_20k2	Strong similarity to rat vanilloid receptor subtype 1.	VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.	Signal transduction
htes3_21d4	Similarity to RCC1-like G exchanging factor RLG	RCC1 is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein.	Signal transduction
htes3_23n19	Similarity to rat protein kinase C-interacting RBCC protein 1	Protein kinase C-interacting protein	Signal transduction
htes3_4f5	Similarity to S.pombe "beta-transducin"	Contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins; in addition, a Cytochrome C family heme-binding site signature is present.	Signal Transduction
htes3_6c11	Strong similarity to YNL132w	Could be a steroid receptor	Signal transduction
htes3_8e24	Related to yeast YGL099w and mouse MMR1 putative GTP-binding proteins.	a novel 658 amino acid putative GTP-binding protein	Signal Transduction
hute1_20g21	Ras inhibitor	Receptor tyrosine kinase (RTK)/RAS/RAF kinase signaling cascade	Signal transduction
hute1_22d2	Similarity to GTP-binding proteins	GTP-binding proteins	Signal transduction

hute1_22e12	Strong similarity to S.cerevisiae YG1054c and cornichon	The Drosophila cni and mammalian proteins cornicon are part of a signal transduction pathway involving hte EGP-receptor	Signal transduction
-------------	---	---	---------------------

Group Metabolism

Clone ID	Homology	Function	Group
hbr2_398	Similarity to N-terminal Acetyltransferase Complex ARD1 homolog	In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1.	Metabolism
hbr2_62o17	Similarity to apolipoprotein E receptor	Low-density lipoprotein (LDL) receptors are the major cholesterol-carrying lipoproteins of plasma. The novel protein contains an additional leucine zipper suitable for protein-protein interaction.	Metabolism
hbr2_6b24	Similar to dUDP-6-deoxy-L-mannose-dehydrogenases	UDP-6-deoxy-L-mannose-dehydrogenase	Metabolism
hbr2_78k24	Similarity to Mus musculus ubiquitin specific protease UBP43.	The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated protein	Metabolism
hbr2_24b15	Similarity to phosphomannomutases	Phosphoserine signature typical for Phosphoglucosyltransferase or Phosphomannomutase ; conversion of Hexose phosphates.	Metabolism
hbr2_30l7	Strong similarity NADH Oxidoreductase B22 subunit-	The new protein is the human orthologue of the bovine EC 1.6.5.3. chain C1-B22 and therefore part of the human respiratory chain.	Metabolism
hbr2_46j20	Strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	The new protein seems to be the human 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.	Metabolism
htea3_15c24	Strong similarity to 2-hydroxyacid dehydrogenases	New human 2-hydroxyacid dehydrogenase	Metabolism
htea3_17l17	Strong similarity to transketolases	Transketolase testis specific	Metabolism
htea3_27d1	Similarity to ubiquitin-specific proteases	Protease	Metabolism
htea3_2a17	Similar to thiol-proteases	Putative thiol-protease	Metabolism
htea3_35b5	Strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A	ATPase	Metabolism
htea3_35k16	Similarity to acyl-CoA synthetase	Acyl-CoA synthetase	Metabolism
htea3_35n12	Strong similarity to ADP/ATP carrier proteins	Involved in mitochondrial energy metabolism	Metabolism
htea3_35n9	Carboxylesterase, splice variant	Carboxylesterase	Metabolism
hute1_20b19	Similarity to sarcosine oxidases	Sarcosine oxidases	Metabolism
hute1_20m24	Strong similarity to S.cerevisiae Alg9p probable mannosyltransferase	Possible mannosyltransferase	Metabolism
hute1_21e13	Strong similarity to heat shock 27K proteins	Heat shock protein related new subtilase	Metabolism

Group transcription factors

CloneID	Homology	Function	Group
hfk2_46k19	Strong similarity to pterin-4-alpha-carbinolamine dehydratase	DcoH is a bifunctional protein, complexed with bioplerin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the bioplerin cofactor of phenylalanine hydroxylase	Transcription factor
hfk2_47a4	Similarity to zinc fingers	New putative transcription factor with one C2H2 zinc fingers.	Transcription factor
htes3_2e12	Similarity to finger proteins	Transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein	Transcription factors
htes3_21j15	3 strong similarity to "NY-CO-33"	Transcription factor	Transcription factors
htes3_17n12	Nearly identical to mouse SOX-LZ	SOX-LZ, related to SRY and HMG-box-Proteins	Transcription factors
hutel_18i19	Similarity to transcription factor SF3	The SREBP-2 protein is cleaved to release soluble NH2-terminal that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis; a lim domain; shows similarity to the common sunflower transcripti	Transcription factor
hutel_1i2	Similarity to Dictostelium myosin heavy chain kinase	Zn-finger protein	Transcription factor

Group uterus associated

CloneID Accession	Unikamology	Function	Group
hutel_17k7	Similarity to HPBRII-4 MRNA	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_18c12	Similarity to candidate tumor suppressor p3ING1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_1814	Weak similarity to C.elegans D2085.2	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_19g19	Partial similarity to bovine elastin fragment	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_19j11	Strong similarity to KIAA0231, similarity to ras binding protein Sur8	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_22n2	Similar to F46P6.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_21d15	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_22o2	Similarity to S.pombe SPBC3E7.03c	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_23g11	Similarity to SPAC31G5.12c and Maf1p	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated

Prosites Key

NAME: N-glycosylation site.
 CONSENSUS: N-{P}-{ST}-{P}.

NAME: Glycosaminoglycan attachment site.
 CONSENSUS: S-G-x-G.

NAME: Tyrosine sulfation site.

NAME: cAMP- and cGMP-dependent protein kinase phosphorylation site.
 CONSENSUS: [RK](2)-x-{ST}.

NAME: Protein kinase C phosphorylation site.
 CONSENSUS: [ST]-x-[RK].

NAME: Casein kinase II phosphorylation site.
 CONSENSUS: [ST]-x(2)-{DE}.

NAME: Tyrosine kinase phosphorylation site.
 CONSENSUS: [RK]-x(2,3)-{DE}-x(2,3)-Y.

NAME: N-myristoylation site.
 CONSENSUS: G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.

NAME: Amidation site.
 CONSENSUS: x-G-[RK]-[RK].

NAME: Aspartic acid and asparagine hydroxylation site.
 CONSENSUS: C-x-[DN]-x(4)-[FY]-x-C-x-C.

NAME: Vitamin K-dependent carboxylation domain.
 CONSENSUS: x(12)-E-x(3)-E-x-C-x(6)-[DEN]-x-[LIVMFY]-x(9)-[FYW].

NAME: Phosphopantetheine attachment site.
 CONSENSUS: [DEQGSTALMKRH]-[LIVMFYSTAC]-[GNQ]-[LIVMFYAG]-[DNEKHS]-S-[LIVMST]-
 CONSENSUS: {PCFY}-[STAGCPQLIVMF]-[LIVMATN]-[DENQGTAKRHLM]-[LIVMWSTA]-[LIVGSTACR]-
 CONSENSUS: x(2)-[LIVMFA].

NAME: Acyl carrier protein phosphopantetheine domain profile.

NAME: Prokaryotic membrane lipoprotein lipid attachment site.
 CONSENSUS: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C.

NAME: Prokaryotic N-terminal methylation site.
 CONSENSUS: [KRHEQSTAG]-G-{FYLVIM}-[ST]-[LT]-[LIVP]-E-[LIVMFWSTAG](14).

NAME: Prenyl group binding site (CAAX box).
 CONSENSUS: C-{DENQ}-[LIVM]-x>.

NAME: Protein splicing signature.
 CONSENSUS: [DNEG]-x-[LIVFA]-[LIVMY]-[LVAST]-H-N-[STC].

NAME: Endoplasmic reticulum targeting sequence.
 CONSENSUS: [KRHQSA]-[DENQ]-E-L>.

NAME: Microbodies C-terminal targeting signal.
 CONSENSUS: [STAGCN]-[RKH]-[LIVMAFY]>.

NAME: Gram-positive cocci surface proteins 'anchoring' hexapeptide.
 CONSENSUS: L-P-x-T-G-[STGAVDE].

NAME: Bipartite nuclear targeting sequence.

NAME: Cell attachment sequence.
 CONSENSUS: R-G-D.

NAME: ATP/GTP-binding site motif A (P-loop).
 CONSENSUS: [AG]-x(4)-G-K-[ST].

NAME: Cyclic nucleotide-binding domain signature 1.
 CONSENSUS: [LIVM]-[VIC]-x(2)-G-[DENQTA]-x-[GAC]-x(2)-[LIVMFY](4)-x(2)-G.

NAME: Cyclic nucleotide-binding domain signature 2.

CONSENSUS: [LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x-[STACV].

NAME: cAMP/cGMP binding motif.

NAME: EF-hand calcium-binding domain.

CONSENSUS: D-x-[DNS]-[ILVFW]-[DENSTG]-[DNQHRK]-[GP]-[LIVMC]-[DENQSTAGC]-x(2)-

CONSENSUS: [DE]-[LIVMFYW].

NAME: Actinin-type actin-binding domain signature 1.

CONSENSUS: [EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N.

NAME: Actinin-type actin-binding domain signature 2.

CONSENSUS: [LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x-[DEAG]-x(4)-

CONSENSUS: [LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x-[LIVM](2).

NAME: Anaphylatoxin domain signature.

CONSENSUS: [CSH]-C-x(2)-[GAP]-x(7,8)-[GASTDEQR]-C-[GASTDEQL]-x(3,9)-[GASTDEQN]-x(2)-

CONSENSUS: [CE]-x(6,7)-C-C.

NAME: Anaphylatoxin domain profile.

NAME: Apple domain.

CONSENSUS: C-x(3)-[LIVMFY]-x(5)-[LIVMFY]-x(3)-[DENQ]-[LIVMFY]-x(10)-C-x(3)-C-T-

CONSENSUS: x(4)-C-x-[LIVMFY]-F-x-[FY]-x(13,14)-C-x-[LIVMFY]-[RK]-x-[ST]-x(14,15)-

CONSENSUS: S-G-x-[ST]-[LIVMFY]-x(2)-C.

NAME: Band 4.1 family domain signature 1.

CONSENSUS: W-[LIV]-x(3)-[KRQ]-x-[LIVM]-x(2)-[QH]-x(0,2)-[LIVMF]-x(6,8)-[LIVMF]-

CONSENSUS: x(3,5)-F-[FY]-x(2)-[DENS].

NAME: Band 4.1 family domain signature 2.

CONSENSUS: [HYW]-x(9)-[DENQSTV]-[SA]-x(3)-[FY]-[LIVM]-x(2)-[ACV]-x(2)-[LM]-x(2)-

CONSENSUS: [FY]-G-x-[DENQST]-[LIVMFYS].

NAME: Band 4.1 family domain profile.

NAME: Clq domain signature.

CONSENSUS: F-x(5)-[ND]-x(4)-[FYWL]-x(6)-F-x(5)-G-x-Y-x-F-x-[FY].

NAME: C-terminal cystine knot signature.

CONSENSUS: C-C-x(13)-C-x(2)-[GN]-x(12)-C-x-C-x(2,4)-C.

NAME: C-terminal cystine knot profile.

NAME: CUB domain profile.

NAME: Death domain profile.

NAME: EGF-like domain signature 1.

CONSENSUS: C-x-C-x(5)-G-x(2)-C.

NAME: EGF-like domain signature 2.

CONSENSUS: C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C.

NAME: Calcium-binding EGF-like domain pattern signature.

CONSENSUS: [DEQN]-x-[DEQN](2)-C-x(3,14)-C-x(3,7)-C-x-[DN]-x(4)-[FY]-x-C.

NAME: Laminin-type EGF-like (LE) domain signature.

CONSENSUS: C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C.

NAME: Coagulation factors 5/8 type C domain (FA58C) signature 1.

CONSENSUS: [GAS]-W-x(7,15)-[FYW]-[LIV]-x-[LIVFA]-[GSTDEN]-x(6)-[LIVF]-x(2)-[IV]-x-

CONSENSUS: [LIVT]-[QKM]-G.

NAME: Coagulation factors 5/8 type C domain (FA58C) signature 2.

CONSENSUS: P-x(8,10)-[LM]-R-x-[GE]-[LIVP]-x-G-C.

NAME: Forkhead-associated (FHA) domain profile.

NAME: Fibrinogen beta and gamma chains C-terminal domain signature.

CONSENSUS: W-W-[LIVMFYW]-x(2)-C-x(2)-[GSA]-x(2)-N-G.

NAME: Type I fibronectin domain.

CONSENSUS: C-x(6,8)-[LFY]-x(5)-[FYW]-x-[RK]-x(8,10)-C-x-C-x(6,9)-C.

NAME: Type II fibronectin collagen-binding domain.

CONSENSUS: C-x(2)-P-F-x-[FYWI]-x(7)-C-x(8,10)-W-C-x(4)-[DNSR]-[FYW]-x(3,5)-[FYW]-x-

CONSENSUS: [FYWI]-C.

NAME: Hemopexin domain signature.

CONSENSUS: [LIFAT]-x(3)-W-x(2,3)-[PE]-x(2)-[LIVMFY]-[DENQS]-[STA]-[AV]-[LIVMFY].

NAME: Kringle domain signature.

CONSENSUS: [FY]-C-R-N-P-[DNR].

NAME: Kringle domain profile.

NAME: LDL-receptor class A (LDLRA) domain signature.

CONSENSUS: C-[VILMA]-x(5)-C-[DNH]-x(3)-[DENQHT]-C-x(3,4)-[STADE]-[DEH]-[DE]-x(1,5)-

CONSENSUS: C.

NAME: LDL-receptor class A (LDLRA) domain profile.

NAME: C-type lectin domain signature.

CONSENSUS: C-[LIVMFYATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-

CONSENSUS: C.

NAME: C-type lectin domain profile.

NAME: Link domain signature.

CONSENSUS: C-x(15)-A-x(3,4)-G-x(3)-C-x(2)-G-x(8,9)-P-x(7)-C.

NAME: Osteonectin domain signature 1.

CONSENSUS: C-x-[DN]-x(2)-C-x(2)-G-[KRH]-x-C-x(6,7)-P-x-C-x-C-x(3,5)-C-P.

NAME: Osteonectin domain signature 2.

CONSENSUS: F-P-x-R-[IM]-x-D-W-L-x-[NQ].

NAME: Somatomedin B domain signature.

CONSENSUS: C-x-C-x(3)-C-x(5)-C-C-x-[DN]-[FY]-x(3)-C.

NAME: Thyroglobulin type-1 repeat signature.

CONSENSUS: [FYWHP]-x-P-x-C-x(3,4)-G-x-[FYW]-x(3)-Q-C-x(4,10)-C-[FYW]-C-V-x(3,4)-

CONSENSUS: [SG].

NAME: P-type 'Trefoil' domain signature.

CONSENSUS: R-x(2)-C-x-[FYPST]-x(3,4)-[ST]-x(3)-C-x(4)-C-C-[FYWH].

NAME: Cellulose-binding domain, bacterial type.

CONSENSUS: W-N-[STAGR]-[STDN]-[LIVM]-x(2)-[GST]-x-[GST]-x(2)-[LIVMFT]-[GA].

NAME: Cellulose-binding domain, fungal type.

CONSENSUS: C-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]-x(2)-Q-C.

NAME: Chitin recognition or binding domain signature.

CONSENSUS: C-x(4,5)-C-C-S-x(2)-G-x-C-G-x(4)-[FYW]-C.

NAME: Barwin domain signature 1.

CONSENSUS: C-G-[KR]-C-L-x-V-x-N.

NAME: Barwin domain signature 2.

CONSENSUS: V-[DN]-Y-[EQ]-F-V-[DN]-C.

NAME: BIR repeat.

CONSENSUS: [HKEPILVY]-x(2)-R-x(3,7)-[FYW]-x(11,14)-[STAN]-G-[LMF]-X-[FYHDA]-X(4)-

CONSENSUS: [DESL]-X(2,3)-C-X(2)-C-X(6)-[WA]-X(9)-H-X(4)-[PRSD]-X-C-X(2)-[LIVMA].

NAME: WAP-type 'four-disulfide core' domain signature.

CONSENSUS: C-x-[C]-[DN]-x(2)-C-x(5)-C-C.

NAME: Phorbol esters / diacylglycerol binding domain.

CONSENSUS: H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)-C-x(2)-C-x(4)-[HD]-

CONSENSUS: x(2)-C-x(5,9)-C.

NAME: C2 domain signature.

CONSENSUS: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D-[PA]-[FY].

NAME: C2-domain profile.

NAME: CAP-Gly domain signature.

CONSENSUS: G-x(8,10)-[FYW]-x-G-[LIVM]-x-[LIVMFY]-x(4)-G-K-[NH]-x-G-[STAR]-x(2)-G-x(2)-[LY]-F.

NAME: Ly-6 / u-PAR domain signature.

CONSENSUS: [EQR]-C-[LIVMFYAH]-x-C-x(5,8)-C-x(3,8)-[EDNQSTV]-C-{C}-x(5)-C-x(12,24)-C.

NAME: MAM domain signature.

CONSENSUS: G-x-[LIVMFY](2)-x(3)-[STA]-x(10,11)-[LV]-x(4)-[LIVMF]-x(6,7)-C-[LIVM]-x-F-x-[LIVMFY]-x(3)-[GSC].

NAME: MAM domain profile.

NAME: PH domain profile.

NAME: Phosphotyrosine interaction domain (PID) profile.

NAME: Src homology 2 (SH2) domain profile.

NAME: Src homology 3 (SH3) domain profile.

NAME: VWFC domain signature.

CONSENSUS: C-x(2,3)-C-x-C-x(6,14)-C-x(3,4)-C-x(2,10)-C-x(9,16)-C-C-x(2,4)-C.

NAME: WW/rsp5/WWP domain signature.

CONSENSUS: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

NAME: WW/rsp5/WWP domain profile.

NAME: ZP domain signature.

CONSENSUS: [LIVMFYW]-x(7)-[STAPDNL]-x(3)-[LIVMFYW]-x-[LIVMFYW]-x-[LIVMFYW]-x(2)-C-
CONSENSUS: [LIVMFYW]-x-[ST]-[PSL]-x(2,4)-[DENS]-x-[STADNQLF]-x(6)-[LIVM](2)-x(3,4)-
CONSENSUS: C.

NAME: S-layer homology domain signature.

CONSENSUS: [LVFYT]-x-[DA]-x(2,5)-[DNQSATPHY]-[WYFPDA]-x(4)-[LIV]-x(2)-[GTALV]-
CONSENSUS: x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x-[PGAV]-x(3,10)-[LIVMA]-
CONSENSUS: [STKR]-[RY]-x-[EQ]-x-[STALIVM].

NAME: 'Homeobox' domain signature.

CONSENSUS: [LIVMFYG]-[ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQESTAIY]-
CONSENSUS: [LIVFSTNKH]-W-[FYVC]-x-[NDQTAH]-x(5)-[RKNALMW].

NAME: 'Homeobox' domain profile.

NAME: 'Homeobox' antennapedia-type protein signature.

CONSENSUS: [LIVMFE]-[FY]-P-W-M-[KRQTA].

NAME: 'Homeobox' engrailed-type protein signature.

CONSENSUS: L-M-A-Q-G-L-Y-N.

NAME: 'Paired box' domain signature.

CONSENSUS: R-P-C-x(11)-C-V-S.

NAME: 'POU' domain signature 1.

CONSENSUS: [RKQ]-R-[LIM]-x-[LF]-G-[LIVMFY]-x-Q-x-[DNQ]-V-G.

NAME: 'POU' domain signature 2.

CONSENSUS: S-Q-[ST]-[TA]-I-[SC]-R-F-E-x-[LSQ]-x-[LI]-[ST].

NAME: Zinc finger, C2H2 type, domain.

CONSENSUS: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H.

NAME: Zinc finger, C3HC4 type (RING finger), signature.

CONSENSUS: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].

NAME: Nuclear hormones receptors DNA-binding region signature.

CONSENSUS: C-x(2)-C-x-[DE]-x(5)-[HN]-[FY]-x(4)-C-x(2)-C-x(2)-F-F-x-R.

NAME: GATA-type zinc finger domain.

CONSENSUS: C-x-[DN]-C-x(4,5)-[ST]-x(2)-W-[HR]-[RK]-x(3)-[GN]-x(3,4)-C-N-[AS]-C.

NAME: Poly(ADP-ribose) polymerase zinc finger domain signature.

CONSENSUS: C-[KR]-x-C-x(3)-I-x-K-x(3)-[RG]-x(16,18)-W-[FYH]-H-x(2)-C.

NAME: Poly(ADP-ribose) polymerase zinc finger domain profile.

NAME: Fungal Zn(2)-Cys(6) binuclear cluster domain signature.

CONSENSUS: [GASTPV]-C-x(2)-C-[RKHSTACW]-x(2)-[RKHQ]-x(2)-C-x(5,12)-C-x(2)-C-x(6,8)-

CONSENSUS: C.

NAME: Fungal Zn(2)-Cys(6) binuclear cluster domain profile.

NAME: Prokaryotic dksA/traR C4-type zinc finger.

CONSENSUS: C-[DES]-x-C-x(3)-I-x(3)-R-x(4)-P-x(4)-C-x(2)-C.

NAME: Copper-fist domain signature.

CONSENSUS: M-[LIVMF](3)-x(3)-K-[MY]-A-C-x(2)-C-I-[KR]-x-H-[KR]-x(3)-C-x-H-x(8)-

CONSENSUS: [KR]-x-[KR]-G-R-P.

NAME: Copper fist DNA binding domain profile.

NAME: Leucine zipper pattern.

CONSENSUS: L-x(6)-L-x(6)-L-x(6)-L.

NAME: bZIP transcription factors basic domain signature.

CONSENSUS: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

NAME: Myb DNA-binding domain repeat signature 1.

CONSENSUS: W-[ST]-x(2)-E-[DE]-x(2)-[LIV].

NAME: Myb DNA-binding domain repeat signature 2.

CONSENSUS: W-x(2)-[LI]-[SAG]-x(4,5)-R-x(8)-[YW]-x(3)-[LIVM].

NAME: Myc-type, 'helix-loop-helix' dimerization domain signature.

CONSENSUS: [DENSTAP]-K-[LIVMWAGSN]-[FYWCPHKR]-[LIVT]-[LIV]-x(2)-[STAV]-[LIVMSTAC]-x-

CONSENSUS: [VMFYH]-[LIVMTA]-{P}-[P]-[LIVMSR].

NAME: p53 tumor antigen signature.

CONSENSUS: M-C-N-S-S-C-M-G-G-M-N-R-R.

NAME: CBF-A/NF- γ B subunit signature.

CONSENSUS: C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C.

NAME: CBF-B/NF- γ A subunit signature.

CONSENSUS: Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E.

NAME: 'Cold-shock' DNA-binding domain signature.

CONSENSUS: [FY]-G-F-I-x(6,7)-[DER]-[LIVM]-F-x-H-x-[STKR]-x-[LIVMFY].

NAME: CTF/NF-I signature.

CONSENSUS: R-K-R-K-Y-F-K-K-H-E-K-R.

NAME: Ets-domain signature 1.

CONSENSUS: L-[FYW]-[QEDH]-F-[LI]-[LVQK]-x-[LI]-L.

NAME: Ets-domain signature 2.

CONSENSUS: [RKH]-x(2)-M-x-Y-[DENQ]-x-[LIVM]-[STAG]-R-[STAG]-[LI]-R-x-Y.

NAME: Ets-domain profile.

NAME: Fork head domain signature 1.

CONSENSUS: [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]-[LIM].

NAME: Fork head domain signature 2.

CONSENSUS: W-[QKR]-[NS]-S-[LIV]-R-H.

NAME: Fork head domain profile.

NAME: HSF-type DNA-binding domain signature.

CONSENSUS: L-x(3)-[FY]-K-H-x-N-x-[STAN]-S-F-[LIVM]-R-Q-L-[NH]-x-Y-x-[FYW]-[RKH]-K-

CONSENSUS: [LIVM].

NAME: Tryptophan pentad repeat (IRF family) signature.

CONSENSUS: W-x-[DNH]-x(5)-[LIVF]-x-[IV]-P-W-x-H-x(9,10)-[DE]-x(2)-[LIVF]-F-[KRQ]-x-

CONSENSUS: [WR]-A.

NAME: LIM domain signature.

CONSENSUS: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)-[LIVMF].

NAME: LIM domain profile.

NAME: NF-kappa-B/Rel/dorsal domain signature.

CONSENSUS: F-R-Y-x-C-E-G.

NAME: MADS-box domain signature.

CONSENSUS: R-x-[RK]-x(5)-I-x-[DN]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)-x(2)-[LIVM]-x-

CONSENSUS: K(2)-A-x-E-[LIVM]-[ST]-x-L-x(4)-[LIVM]-x-[LIVM](3)-x(6)-[LIVMF]-x(2)-

CONSENSUS: [FY].

NAME: MADS-box domain profile.

NAME: T-box domain signature 1.

CONSENSUS: L-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ].

NAME: T-box domain signature 2.

CONSENSUS: [LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x-F.

NAME: TEA domain signature.

CONSENSUS: G-R-N-E-L-I-x(2)-Y-I-x(3)-[TC]-x(3)-R-T-[RK](2)-Q-[LIVM]-S-S-H-[LIVM]-

CONSENSUS: Q-V.

NAME: Transcription factor TFIIIB repeat signature.

CONSENSUS: G-[KR]-x(3)-[STAGN]-x-[LIVMYA]-[GSTA](2)-[CSAV]-[LIVM]-[LIVMFY]-[LIVMA]-

CONSENSUS: [GSA]-[STAC].

NAME: Transcription factor TFIID repeat signature.

CONSENSUS: Y-x-P-x(2)-[IF]-x(2)-[LIVM](2)-x-[KRH]-x(3)-P-[RKQ]-x(3)-L-[LIVM]-F-x-

CONSENSUS: [STN]-G-[KR]-[LIVM]-x(3)-G-[TAGL]-[KR]-x(7)-[AGC]-x(7)-[LIVM].

NAME: TFIIIS zinc ribbon domain signature.

CONSENSUS: C-x(2)-C-x(9)-[LIVMQSAR]-[QH]-[STQL]-[RA]-[SACR]-x-[DE]-[DET]-[PGSEA]-

CONSENSUS: x(6)-C-x(2,5)-C-x(3)-[FW].

NAME: TSC-22 / dip / bun family signature.

CONSENSUS: M-D-L-V-K-x-H-L-x(2)-A-V-R-E-E-V-E.

NAME: Prokaryotic transcription elongation factors signature 1.

CONSENSUS: [ST]-x(2)-[GS]-x(3)-[LI]-x(2)-E-L-x(2)-L-x(3,4)-R-x(2)-[IV]-x(3)-[LIV]-

CONSENSUS: x(6)-G-D-x(2)-E-N-[GSA]-x-Y.

NAME: Prokaryotic transcription elongation factors signature 2.

CONSENSUS: S-x(2)-S-P-[LIVM]-[AG]-x-[SAG]-[LIVM]-[LIVMY]-x(4)-[DG]-[DE].

NAME: DEAD-box subfamily ATP-dependent helicases signature.

CONSENSUS: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN].

NAME: DEAH-box subfamily ATP-dependent helicases signature.

CONSENSUS: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].

NAME: Eukaryotic putative RNA-binding region RNP-1 signature.

CONSENSUS: [RK]-G-[EDRKHPCG]-[AGSCI]-[FY]-[LIVA]-x-[FYLM].

NAME: Fibrillarin signature.

CONSENSUS: [GST]-[LIVMAP]-V-Y-A-[IV]-E-[FY]-[SA]-x-R-x(2)-R-[DE].

NAME: MCM family signature.

CONSENSUS: G-[IVT]-[LVAC](2)-[IVT]-D-[DE]-[FL]-[DNST].

NAME: MCM family domain.

NAME: XPA protein signature 1.

CONSENSUS: C-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C.

NAME: XPA protein signature 2.

CONSENSUS: [LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE].

NAME: XPG protein signature 1.

CONSENSUS: [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K.

NAME: XPG protein signature 2.

CONSENSUS: [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]-[QS]-[CLM].

NAME: Bacterial regulatory proteins, araC family signature.

CONSENSUS: [KRQ]-[LIVMA]-x(2)-[GSTALIV]-[FYWPGDN]-x(2)-[LIVMSA]-x(4,9)-[LIVMF]-

CONSENSUS: x(2)-[LIVMSTA]-[GSTACIL]-x(3)-[GANQRF]-[LIVMFY]-x(4,5)-[LFY]-x(3)-

CONSENSUS: [FYIVA]-[FYWHCM]-x(3)-[GSADENQKR]-x-[NSTAPKL]-[PARL].

NAME: Bacterial regulatory proteins, araC family DNA-binding domain profile.

NAME: Bacterial regulatory proteins, arsR family signature.

CONSENSUS: C-x(2)-D-[LIVM]-x(6)-[ST]-x(4)-S-[HYR]-[HQ].

NAME: Bacterial regulatory proteins, asnC family signature.

CONSENSUS: [GSTAP]-x(2)-[DNEA]-[LIVM]-[GSA]-x(2)-[LIVMFY]-[GN]-[LIVMST]-[ST]-x(6)-R-

CONSENSUS: [LVT]-x(2)-[LIVM]-x(3)-G.

NAME: Bacterial regulatory proteins, crp family signature.

CONSENSUS: [LIVM]-[STAG]-[RHNW]-x(2)-[LIM]-[GA]-x-[LIVMFYA]-[LIVSC]-[GA]-x-[STACN]-

CONSENSUS: x(2)-[MST]-x-[GSTN]-R-x-[LIVMF]-x(2)-[LIVMF].

NAME: Bacterial regulatory proteins, deoR family signature.

CONSENSUS: R-x(3)-[LIVM]-x(3)-[LIVM]-x(16,17)-[STA]-x(2)-T-[LIVMA]-[RH]-[KRNA]-D-

CONSENSUS: [LIVMF].

NAME: Bacterial regulatory proteins, gntR family signature.

CONSENSUS: [LIVAPKR]-[PILV]-x-[EQTIVMR]-x(2)-[LIVM]-x(3)-[LIVMFYK]-x-[LIVFT]-

CONSENSUS: [DNGSTK]-[RGTLV]-x-[STAIVP]-[LIVA]-x(2)-[STAGV]-[LIVMFYH]-x(2)-[LMA].

NAME: Bacterial regulatory proteins, iclR family signature.

CONSENSUS: [GA]-x(3)-[DS]-x(2)-E-x(6)-[CSA]-[LIVM]-[GSA]-x(2)-[LIVM]-[FYH]-[DN].

NAME: Bacterial regulatory proteins, lacI family signature.

CONSENSUS: [LIVM]-x-[DE]-[LIVM]-A-x(2)-[STAGV]-x-V-[GSTP]-x(2)-[STAG]-[LIVMA]-x(2)-

CONSENSUS: [LIVMFYAN]-[LIVMC].

NAME: Bacterial regulatory proteins, luxR family signature.

CONSENSUS: [GDC]-x(2)-[NSTAVY]-x(2)-[IV]-[GSTA]-x(2)-[LIVMFYWCT]-x-[LIVMFYWCR]-x(3)-

CONSENSUS: [NST]-[LIVM]-x(5)-[NRHSA]-[LIVMSTA]-x(2)-[KR].

NAME: Bacterial regulatory proteins, lysR family signature.

CONSENSUS: [NQKRHSTAG]-[LIVMFYTA]-x(2)-[STAGLV]-[STAG]-x(4)-[LIVMYCTQR]-[PSTANLVER]-

CONSENSUS: x-[PSTAGQV]-[PSTAGNVMF]-[LIVMFA]-[STAGH]-x(2)-[LIVMF]-x(2)-[LIVMFW]-

CONSENSUS: [RKEAV]-x(2)-[LIVMFYNTAE]-x(3)-[LIMVT].

NAME: Bacterial regulatory proteins, marR family signature.

CONSENSUS: [STNA]-[LIA]-x-[RNGS]-x(4)-[LM]-[EIV]-x(2)-[GES]-[LFYW]-[LIVC]-x(7)-

CONSENSUS: [DN]-[RKQG]-[RK]-x(6)-T-x(2)-[GA].

NAME: Bacterial regulatory proteins, merR family signature.

CONSENSUS: [GSA]-x-[LIVMFA]-[ASM]-x(2)-[STACLIV]-[GSDENQR]-[LIVC]-[STANHK]-x(3)-

CONSENSUS: [LIVM]-[RHF]-x-[YW]-[DEQ]-x(2,3)-[GHDNQ]-[LIVMF](2).

NAME: Bacterial regulatory proteins, tetR family signature.

CONSENSUS: G-[LIVMFYS]-x(2,3)-[TS]-[LIVMT]-x(2)-[LIVM]-x(5)-[LIVQS]-[STAGENQH]-x-

CONSENSUS: [GPAR]-x-[LIVMF]-[FYST]-x-[HFY]-[FV]-x-[DNST]-K-x(2)-[LIVM].

NAME: Transcriptional antiterminators bglG family signature.

CONSENSUS: [ST]-x-H-x(2)-[FA](2)-[LIVM]-[EQK]-R-x(2)-[QNK].

NAME: Sigma-54 factors family signature 1.

CONSENSUS: P-[LIVM]-x-[LIVM]-x(2)-[LIVM]-A-x(2)-[LIVMF]-x(2)-[HS]-x-S-T-[LIVM]-S-R.

NAME: Sigma-54 factors family signature 2.

CONSENSUS: R-R-T-[IV]-[AT]-K-Y-R.

NAME: Sigma-54 factors family profile.

NAME: Sigma-70 factors family signature 1.

CONSENSUS: [DE]-[LIVMF](2)-[HEQS]-x-G-x-[LIVMFA]-G-L-[LIVMFYE]-x-[GSAM]-[LIVMAP].

NAME: Sigma-70 factors family signature 2.

CONSENSUS: [STN]-x(2)-[DEQ]-[LIVM]-[GAS]-x(4)-[LIVMF]-[PSTG]-x(3)-[LIVMA]-x-[NQR]-

CONSENSUS: [LIVMA]-[EQH]-x(3)-[LIVMFW]-x(2)-[LIVM].

NAME: Sigma-70 factors ECF subfamily signature.

CONSENSUS: [STAIV]-[PQDEL]-[DE]-[LIV]-[LIVTA]-Q-x-[STAV]-[LIVMFYC]-[LIVMAK]-x-

CONSENSUS: [GSTAIV]-[LIMFYWQ]-x(12,14)-[STAP]-[FYW]-[LIF]-x(2)-[IV].

NAME: Sigma-54 interaction domain ATP-binding region A signature.

CONSENSUS: [LIVMFY](3)-x-G-[DEQ]-[STE]-G-[STAV]-G-K-x(2)-[LIVMFY].

NAME: Sigma-54 interaction domain ATP-binding region B signature.

CONSENSUS: [GS]-x-[LIVMF]-x(2)-A-[DNEQASH]-[GNEK]-G-[STIM]-[LIVMFY](3)-[DE]-[EK]-

CONSENSUS: [LIVM].

NAME: Sigma-54 interaction domain C-terminal part signature.

CONSENSUS: [FYW]-P-[GS]-N-[LIVM]-R-[EQ]-L-x-[NHAT].

NAME: Sigma-54 interaction domain profile.

NAME: Single-strand binding protein family signature 1.

CONSENSUS: [LIVMF]-[NST]-[KRT]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVM]-[GST]-x-[DET].

NAME: Single-strand binding protein family signature 2.

CONSENSUS: T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR].

NAME: Bacterial histone-like DNA-binding proteins signature.

CONSENSUS: [GSK]-F-x(2)-[LIVMF]-x(4)-[RKEQA]-x(2)-[RST]-x-[GA]-x-[KN]-P-x-T.

NAME: Dps protein family signature 1.

CONSENSUS: H-[FW]-x-[LIVM]-x-G-x(5)-[LV]-H-x(3)-[DE].

NAME: Dps protein family signature 2.

CONSENSUS: [LIVMFY]-[DH]-x-[LIVM]-[GA]-E-R-x(3)-[LIF]-[GDN]-x(2)-[PA].

NAME: DNA repair protein radC family signature.

CONSENSUS: H-N-H-P-S-G.

NAME: recA signature.

CONSENSUS: A-L-[KR]-[IF]-[FY]-[STA]-[STAD]-[LIVMQ]-R.

NAME: RecF protein signature 1.

CONSENSUS: P-[ED]-x(3)-[LIVM](2)-x-G-[GSAD]-P-x(2)-R-R-x-[FY]-[LIVM]-D.

NAME: RecF protein signature 2.

CONSENSUS: [LIVMFY](2)-x-D-x(2,3)-[SA]-[EH]-L-D-x(2)-[KRH]-x(3)-L.

NAME: RecR protein signature.

CONSENSUS: C-x(2)-C-x(3)-[ST]-x(4)-C-x-I-C-x(4)-R.

NAME: Histone H2A signature.

CONSENSUS: [AC]-G-L-x-F-P-V.

NAME: Histone H2B signature.

CONSENSUS: [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L-x-[KR]-H-A-

CONSENSUS: [LIVM]-[STA]-E-G.

NAME: Histone H3 signature 1.

CONSENSUS: K-A-P-R-K-Q-L.

NAME: Histone H3 signature 2.

CONSENSUS: P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV].

NAME: Histone H4 signature.

CONSENSUS: G-A-K-R-H.

NAME: HMG1/2 signature.

CONSENSUS: [FI]-S-[KR]-K-C-S-[EK]-R-W-K-T-M.

NAME: HMG-I and HMG-Y DNA-binding domain (A+T-hook).

CONSENSUS: [AT]-x(1,2)-[RK](2)-[GP]-R-G-R-P-[RK]-x.

NAME: HMG14 and HMG17 signature.

CONSENSUS: R-R-S-A-R-L-S-A-[RK]-P.

NAME: Bromodomain signature.

CONSENSUS: [STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-
 CONSENSUS: [LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY].

NAME: Bromodomain profile.

NAME: Chromo domain signature.

CONSENSUS: [FYL]-x-[LIVMC]-[KR]-W-x-[GDNR]-[FYWLE]-x(5,6)-[ST]-W-[ES]-[PSTDN]-x(3)-
 CONSENSUS: [LIVMC].

NAME: Chromo and chromo shadow domain profile.

NAME: Regulator of chromosome condensation (RCC1) signature 1.

CONSENSUS: G-x-N-D-x(2)-[AV]-L-G-R-x-T.

NAME: Regulator of chromosome condensation (RCC1) signature 2.

CONSENSUS: [LIVMFA]-[STAGC](2)-G-x(2)-H-[STAGLI]-[LIVMFA]-x-[LIVM].

NAME: Protamine P1 signature.

CONSENSUS: [AV]-R-[NFY]-R-x(2,3)-[ST]-x-S-x-S.

NAME: Nuclear transition protein 1 signature.

CONSENSUS: S-K-R-K-Y-R-K.

NAME: Nuclear transition protein 2 signature 1.

CONSENSUS: H-x(3)-H-S-[NS]-S-x-P-Q-S.

NAME: Nuclear transition protein 2 signature 2.

CONSENSUS: K-x-R-K-x(2)-E-G-K-x(2)-K-[KR]-K.

NAME: Ribosomal protein L1 signature.

CONSENSUS: [IM]-x(2)-[LIVA]-x(2,3)-[LIVM]-G-x(2)-[LMS]-[GSNH]-[PTKR]-[KRAV]-G-x-
 CONSENSUS: [LMF]-P-[DENSTK].

NAME: Ribosomal protein L2 signature.

CONSENSUS: P-x(2)-R-G-[STAIV](2)-x-N-[APK]-x-[DE].

NAME: Ribosomal protein L3 signature.

CONSENSUS: [FL]-x(6)-[DN]-x(2)-[AGS]-x-[ST]-x-G-[KRH]-G-x(2)-G-x(3)-R.

NAME: Ribosomal protein L5 signature.

CONSENSUS: [LIVM]-x(2)-[LIVM]-[STAC]-[GE]-[QV]-x(2)-[LIVMA]-x-[STC]-x-[STAG]-[KR]-
 CONSENSUS: x-[STA].

NAME: Ribosomal protein L6 signature 1.

CONSENSUS: [PS]-[DENS]-x-Y-K-[GA]-K-G-[LIVM].

NAME: Ribosomal protein L6 signature 2.

CONSENSUS: Q-x(3)-[LIVM]-x(2)-[KR]-x(2)-R-x-F-x-D-G-[LIVM]-Y-[LIVM]-x(2)-[KR].

NAME: Ribosomal protein L9 signature.

CONSENSUS: G-x(2)-[GN]-x(4)-V-x(2)-G-[FY]-x(2)-N-[FY]-L-x(5)-[GA]-x(3)-[STN].

NAME: Ribosomal protein L10 signature.

CONSENSUS: [DEH]-x(2)-[GS]-[LIVMF]-[STN]-[VA]-x-[DEQK]-[LIVMA]-x(2)-[LIM]-R.

NAME: Ribosomal protein L11 signature.

CONSENSUS: [RKN]-x-[LIVM]-x-G-[ST]-x(2)-[SNQ]-[LIVM]-G-x(2)-[LIVM]-x(0,1)-[DENG].

NAME: Ribosomal protein L13 signature.

CONSENSUS: [LIVM]-[KRV]-[GK]-M-[LIV]-[PS]-x(4,5)-[GS]-[NQEKRA]-x(5)-[LIVM]-x-[AIV]-
 CONSENSUS: [LFY]-x-[GDN].

NAME: Ribosomal protein L14 signature.

CONSENSUS: [GA]-[LIV](3)-x(9,10)-[DNS]-G-x(4)-[FY]-x(2)-[NT]-x(2)-V-[LIV].

NAME: Ribosomal protein L15 signature.

CONSENSUS: K-[LIVM](2)-[GAL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x-[LIVMF]-x(3,4)-
 CONSENSUS: [LIVMFC]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G.

NAME: Ribosomal protein L16 signature 1.

CONSENSUS: [KR]-R-x-[GSAC]-[KQVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP].

NAME: Ribosomal protein L16 signature 2.

CONSENSUS: R-M-G-x-[GR]-K-G-x(4)-[FWKR].

NAME: Ribosomal protein L17 signature.

CONSENSUS: I-x-[ST]-[GT]-x(2)-[KR]-x-K-x(6)-[DE]-x-[LIMV]-[LIVMT]-T-x-[STAG]-[KR].

NAME: Ribosomal protein L19 signature.

CONSENSUS: [RT]-[KRSVY]-[GSA]-x-V-[RS]-[KR]-[SA]-K-L-Y-Y-L-R.

NAME: Ribosomal protein L20 signature.

CONSENSUS: K-x(3)-[KRC]-x-[LIVM]-W-[IV]-[STNALV]-R-[LIVM]-N-x(3)-[RKH].

NAME: Ribosomal protein L21 signature.

CONSENSUS: [IVT]-x(3)-[KR]-x(3)-[KRQ]-K-x(6)-G-[HF]-R-[RQ]-x(2)-T.

NAME: Ribosomal protein L22 signature.

CONSENSUS: [RKQN]-x(4)-[RH]-[GAS]-x-G-[KRQS]-x(9)-[HDN]-[LIVM]-x-[LIVMS]-x-[LIVM].

NAME: Ribosomal protein L23 signature.

CONSENSUS: [RK](2)-[AM]-[IVFYT]-[IV]-[RKT]-L-[STANQK]-x(7)-[LIVMFT].

NAME: Ribosomal protein L24 signature.

CONSENSUS: [GDEN]-D-x-V-x-[IV]-[LIVMA]-x-G-x(2)-[KA]-[GN]-x(2,3)-[GA]-x-[IV].

NAME: Ribosomal protein L27 signature.

CONSENSUS: G-x-[LIVM](2)-x-R-Q-R-G-x(5)-G.

NAME: Ribosomal protein L29 signature.

CONSENSUS: [KNQS]-[PSTL]-x(2)-[LIMFA]-[KRGSA]-x-[LIVYSTA]-[KR]-[KRH]-[DESTANRL].

CONSENSUS: [LIV]-A-[KRCQVT]-[LIVMA].

NAME: Ribosomal protein L30 signature.

CONSENSUS: [IVT]-[LIVM]-x(2)-[LF]-x-[LI]-x-[KRHQEG]-x(2)-[STNQH]-x-[IVT].

CONSENSUS: x(10)-[LMS]-[LIV]-x(2)-[LIVA]-x(2)-[LMFY]-[IVT].

NAME: Ribosomal protein L31 signature.

CONSENSUS: H-P-F-[FY]-[TI]-x(9)-G-R-[AV]-x-[KR].

NAME: Ribosomal protein L33 signature.

CONSENSUS: Y-x-[ST]-x-[KR]-[NS]-x(4)-[PAT]-x(1,2)-[LIVM]-[EA]-x(2)-K-[FY]-[CSD].

NAME: Ribosomal protein L34 signature.

CONSENSUS: K-[RG]-T-[FYWL]-[EQS]-x(5)-[KRHS]-x(4,5)-G-F-x(2)-R.

NAME: Ribosomal protein L35 signature.

CONSENSUS: [LIVM]-K-[TV]-x(2)-[GSA]-[SAIL]-x-K-R-[LIVMFY]-[KRL].

NAME: Ribosomal protein L36 signature.

CONSENSUS: C-x(2)-C-x(2)-[LIVM]-x-R-x(3)-[LIVMN]-x-[LIVM]-x-C-x(3,4)-[KR]-H-x-Q-x-Q.

NAME: Ribosomal protein L1e signature.

CONSENSUS: N-x(3)-[KR]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-[ST]-[SGA]-x(7)-[RK]-G-H.

NAME: Ribosomal protein L6e signature.

CONSENSUS: N-x(2)-P-L-R-R-x(4)-[FY]-V-I-A-T-S-x-K.

NAME: Ribosomal protein L7Ac signature.

CONSENSUS: [CA]-x(4)-[IV]-P-[FY]-x(2)-[LIVM]-x-[GSQ]-[KRQ]-x(2)-L-G.

NAME: Ribosomal protein L10e signature.

CONSENSUS: R-x-A-[FYW]-G-K-[PA]-x-G-x(2)-A-R-V.

NAME: Ribosomal protein L13e signature.

CONSENSUS: [KR]-Y-x(2)-K-[LIVM]-R-[STA]-G-[KR]-G-F-[ST]-L-x-E.

NAME: Ribosomal protein L15e signature.

CONSENSUS: [DE]-[KR]-A-R-x-L-G-[FY]-x-[SAP]-x(2)-G-[LIVMFY](4)-R-x-R-V-x-R-G.

NAME: Ribosomal protein L18e signature.

CONSENSUS: [KRE]-x-L-x(2)-[PS]-[KR]-x(2)-[RH]-[PSA]-x-[LIVM]-[NS]-[LIVM]-x-[RK].

CONSENSUS: [LIVM].

NAME: Ribosomal protein L19e signature.

CONSENSUS: R-x-[KR]-x(5)-[KR]-x(3)-[KRH]-x(2)-G-x-G-x-R-x-G-x(3)-A-R-x(3)-[KQ].

CONSENSUS: x(2)-W-x(7)-R-x(2)-L-x(3)-R.

NAME: Ribosomal protein L21e signature.
 CONSENSUS: G-[DE]-x-V-x(10)-[GV]-x(2)-[FYH]-x(2)-[FY]-x-G-x-T-G.

NAME: Ribosomal protein L24e signature.
 CONSENSUS: [FY]-x-[GS]-x(2)-[IV]-x-P-G-x-G-x(2)-[FYV]-x-[KRHE]-x-D.

NAME: Ribosomal protein L27e signature.
 CONSENSUS: G-K-N-x-W-F-F-x-K-L-R-F>.

NAME: Ribosomal protein L30e signature 1.
 CONSENSUS: [STA]-x(5)-G-x-[QKR]-x(2)-[LIVM]-[KQT]-x(2)-[KR]-x-G-x(2)-K-x-[LIVM](3).

NAME: Ribosomal protein L30e signature 2.
 CONSENSUS: [DE]-L-G-[STA]-x(2)-G-[KR]-x(6)-[LIVM]-x-[LIVM]-x-[DEN]-x-G.

NAME: Ribosomal protein L31e signature.
 CONSENSUS: V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AK]-x-W-x-[KR]-G.

NAME: Ribosomal protein L32e signature.
 CONSENSUS: F-x-R-x(4)-[KR]-x(2)-[KR]-[LIVM]-x(3)-W-R-[KR]-x(2)-G.

NAME: Ribosomal protein L34e signature.
 CONSENSUS: Y-x-[ST]-x-S-[NY]-x(5)-[KR]-T-P-G.

NAME: Ribosomal protein L35Ae signature.
 CONSENSUS: G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P.

NAME: Ribosomal protein L36e signature.
 CONSENSUS: P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR].

NAME: Ribosomal protein L37e signature.
 CONSENSUS: G-T-x-[SA]-x-G-x-[KR]-x(3)-[ST]-x(0,1)-H-x(2)-C-x-R-C-G.

NAME: Ribosomal protein L39e signature.
 CONSENSUS: [KRA]-T-x(3)-[LIVM]-[KRQF]-x-[NHS]-x(3)-R-[NHY]-W-R-R.

NAME: Ribosomal protein L44e signature.
 CONSENSUS: K-x-[TV]-K-K-x(2)-L-[KR]-x(2)-C.

NAME: Ribosomal protein S2 signature 1.
 CONSENSUS: [LIVMFA]-x(2)-[LIVMFYC](2)-x-[STAC]-[GSTANQEK]-[STALV]-[HY]-[LIVMF]-G.

NAME: Ribosomal protein S2 signature 2.
 CONSENSUS: P-x(2)-[LIVMF](2)-[LIVMS]-x-[GDN]-x(3)-[DENL]-x(3)-[LIVM]-x-E-x(4)-
 CONSENSUS: [GNQKRH]-[LIVM]-[AP].

NAME: Ribosomal protein S3 signature.
 CONSENSUS: [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-
 CONSENSUS: [DENQ]-x(7)-[LMT]-x(2)-G-x(2)-G.

NAME: Ribosomal protein S4 signature.
 CONSENSUS: [LIVM]-[DE]-x-R-L-x(3)-[LIVMC]-[VMFYHQ]-[KRT]-x(3)-[STAGCF]-x-[ST]-x(3)-
 CONSENSUS: [SAI]-[KR]-x-[LIVMF](2).

NAME: Ribosomal protein S5 signature.
 CONSENSUS: G-[KRQ]-x(3)-[FY]-x-[ACV]-x(2)-[LIVMA]-[LIVM]-[AG]-[DN]-x(2)-G-x-
 CONSENSUS: [LIVM]-G-x-[SAG]-x(5,6)-[DEQ]-[LIVM]-x(2)-A-[LIVMF].

NAME: Ribosomal protein S6 signature.
 CONSENSUS: G-x-[KRC]-[DENQRH]-L-[SA]-Y-x-I-[KRNSA].

NAME: Ribosomal protein S7 signature.
 CONSENSUS: [DENSK]-x-[LIVMET]-x(3)-[LIVMFT](2)-x(6)-G-K-[KR]-x(5)-[LIVMF]-[LIVMFC]-
 CONSENSUS: x(2)-[STA].

NAME: Ribosomal protein S8 signature.
 CONSENSUS: [GE]-x(2)-[LIV](2)-[STY]-T-x(2)-G-[LIVM](2)-x(4)-[AG]-[KRHAYT].

NAME: Ribosomal protein S9 signature.
 CONSENSUS: G-G-G-x(2)-[GSA]-Q-x(2)-[SA]-x(3)-[GSA]-x-[GSTAV]-[KR]-[GSAL]-[LIF].

NAME: Ribosomal protein S10 signature.
 CONSENSUS: [AV]-x(3)-[GDNSR]-[LIVMSTA]-x(3)-G-P-[LIVM]-x-[LIVM]-P-T.

NAME: Ribosomal protein S11 signature.
 CONSENSUS: [LIVMF]-x-[GSTAC]-[LIVMF]-x(2)-[GSTAL]-x(0,1)-[GSN]-[LIVMF]-x-[LIVM]-
 CONSENSUS: x(4)-[DEN]-x-T-P-x-[PA]-[STCH]-[DN].

NAME: Ribosomal protein S12 signature.
 CONSENSUS: [RK]-x-P-N-S-[AR]-x-R.

NAME: Ribosomal protein S13 signature.
 CONSENSUS: [KRQS]-G-x-R-H-x(2)-[GSNH]-x(2)-[LIVMC]-R-G-Q.

NAME: Ribosomal protein S14 signature.
 CONSENSUS: [RP]-x(0,1)-C-x(11,12)-[LIVMF]-x-[LIVMF]-[SC]-[RG]-x(3)-[RN].

NAME: Ribosomal protein S15 signature.
 CONSENSUS: [LIVM]-x(2)-H-[LIVMFY]-x(5)-D-x(2)-[SAGN]-x(3)-[LF]-x(9)-[LIVM]-x(2)-
 CONSENSUS: [FY].

NAME: Ribosomal protein S16 signature.
 CONSENSUS: [LIVMT]-x-[LIVM]-[KR]-L-[STAK]-R-x-G-[AKR].

NAME: Ribosomal protein S17 signature.
 CONSENSUS: G-D-x-[LIV]-x-[LIVA]-x-[QEK]-x-[RK]-P-[LIV]-S.

NAME: Ribosomal protein S18 signature.
 CONSENSUS: [IV]-[DY]-Y-x(2)-[LIVMT]-x(2)-[LIVM]-x(2)-[FYT]-[LIVM]-[ST]-[DERP]-x-
 CONSENSUS: [GY]-K-[LIVM]-x(3)-R-[LIVMAS].

NAME: Ribosomal protein S19 signature.
 CONSENSUS: [STDNQ]-G-[KRQM]-x(6)-[LIVM]-x(4)-[LIVM]-[GSD]-x(2)-[LF]-[GAS]-[DE]-F-
 CONSENSUS: x(2)-[ST].

NAME: Ribosomal protein S21 signature.
 CONSENSUS: [DE]-x-A-[LY]-[KR]-R-F-K-[KR]-x(3)-[KR].

NAME: Ribosomal protein S3Ae signature.
 CONSENSUS: [LIV]-x-[GH]-R-[IV]-x-E-x-[SC]-L-x-D-L.

NAME: Ribosomal protein S4e signature.
 CONSENSUS: H-x-K-R-[LIVM]-[SAN]-x-P-x(2)-W-x-[LIVM]-x-[KR].

NAME: Ribosomal protein S6e signature.
 CONSENSUS: [LIVM]-[STAMR]-G-G-x-D-x(2)-G-x-P-M.

NAME: Ribosomal protein S7e signature.
 CONSENSUS: [KR]-L-x-R-E-L-E-K-K-F-[SAP]-x-[KR]-H.

NAME: Ribosomal protein S8e signature.
 CONSENSUS: R-x(2)-T-G-[GA]-x(5)-[HR]-K-[KR]-x-K-x-E-[LM]-G.

NAME: Ribosomal protein S12e signature.
 CONSENSUS: A-L-[KRQP]-x-V-L-x(2)-[SA]-x(3)-[DN]-G-L.

NAME: Ribosomal protein S17e signature.
 CONSENSUS: A-x-I-x-[ST]-K-x-L-R-N-[KR]-I-A-G-[FY]-x-T-H.

NAME: Ribosomal protein S19e signature.
 CONSENSUS: P-x(6)-[SAN]-x(2)-[LIVMA]-x-R-x-[ALIV]-[LV]-Q-x-L-[EQ].

NAME: Ribosomal protein S21e signature.
 CONSENSUS: L-Y-V-P-R-K-C-S-[SA].

NAME: Ribosomal protein S24e signature.
 CONSENSUS: [FA]-G-x(2)-[KR]-[STA]-x-G-[FY]-[GA]-x-[LIVM]-Y-[DN]-[SN].

NAME: Ribosomal protein S26e signature.
 CONSENSUS: [YH]-C-V-S-C-A-I-H.

NAME: Ribosomal protein S27e signature.
 CONSENSUS: [QK]-C-x(2)-C-x(6)-F-[GS]-x-[PSA]-x(5)-C-x(2)-C-[GS]-x(2)-L-x(2)-P-x-G.

NAME: Ribosomal protein S28e signature.
 CONSENSUS: E-[ST]-E-R-E-A-R-x-L.

NAME: DNA mismatch repair proteins mutL / hexB / PMS1 signature.

CONSENSUS: G-F-R-G-E-A-L.

NAME: DNA mismatch repair proteins mutS family signature.

CONSENSUS: [ST]-[LIVM]-x-[LIVM]-x-D-E-[LIVMY]-[GC]-[RKH]-G-[GST]-x(4)-G.

NAME: mutT domain signature.

CONSENSUS: G-x(5)-E-x(4)-[STAGC]-[LIVMAC]-x-R-E-[LIVMFT]-x-E-E.

NAME: DnaA protein signature.

CONSENSUS: I-[GA]-x(2)-[LIVMF]-[SGDNK]-x(0,1)-[KR]-x-H-[STP]-[STV]-[LIVM](2)-x-

CONSENSUS: [SA]-x(2)-[KRE]-[LIVM].

NAME: Small, acid-soluble spore proteins, alpha/beta type, signature 1.

CONSENSUS: K-x-E-[LIV]-A-x-[DE]-[LIVMF]-G-[LIVMF].

NAME: Small, acid-soluble spore proteins, alpha/beta type, signature 2.

CONSENSUS: [KR]-[SAQ]-x-G-x-V-G-G-x-[LIVM]-x-[KR](2)-[LIVM](2).

NAME: Zinc-containing alcohol dehydrogenases signature.

CONSENSUS: G-H-E-x(2)-G-x(5)-[GA]-x(2)-[IVSAC].

NAME: Quinone oxidoreductase / zeta-crystallin signature.

CONSENSUS: [GSD]-[DEQH]-x(2)-L-x(3)-[SA](2)-G-G-x-G-x(4)-Q-x(2)-[KR].

NAME: Iron-containing alcohol dehydrogenases signature 1.

CONSENSUS: [STALIV]-[LIVF]-x-[DE]-x(6,7)-P-x(4)-[ALIV]-x-[GST]-x(2)-D-[TAIVM]-

CONSENSUS: [LIVMF]-x(4)-E.

NAME: Iron-containing alcohol dehydrogenases signature 2.

CONSENSUS: [GSW]-x-[LIVTSACD]-[GH]-x(2)-[GSAE]-[GSHYQ]-x-[LIVTP]-[GAST]-[GAS]-x(3)-

CONSENSUS: [LIVMT]-x-[HNS]-[GA]-x-[GTAC].

NAME: Short-chain dehydrogenases/reductases family signature.

CONSENSUS: [LIVSPADNK]-x(12)-Y-[PSTAGNCV]-[STAGNQCIVM]-[STAGC]-K-[PC]-[SAGFR]-

CONSENSUS: [LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)-[LIVMFYWGAPTHQ]-[GSACQRHM].

NAME: Aldo/keto reductase family signature 1.

CONSENSUS: G-[FY]-R-[HSAL]-[LIVMF]-D-[STAGC]-[AS]-x(5)-E-x(2)-[LIVM]-G.

NAME: Aldo/keto reductase family signature 2.

CONSENSUS: [LIVMFY]-x(9)-[KREQ]-x-[LIVM]-G-[LIVM]-[SC]-N-[FY].

NAME: Aldo/keto reductase family putative active site signature.

CONSENSUS: [LIVM]-[PAIV]-[KR]-[ST]-x(4)-R-x(2)-[GSTAEQK]-[NSL]-x(2)-[LIVMFA].

NAME: Homoserine dehydrogenase signature.

CONSENSUS: A-x(3)-G-[LIVMFY]-[STAG]-x(2,3)-[DNS]-P-x(2)-D-[LIVM]-x-G-x-D-x(3)-K.

NAME: NAD-dependent glycerol-3-phosphate dehydrogenase signature.

CONSENSUS: G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x-[DE]-G-[LIVM]-x-

CONSENSUS: [LIVMFYW]-G-x-N.

NAME: FAD-dependent glycerol-3-phosphate dehydrogenase signature 1.

CONSENSUS: [TV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G.

NAME: FAD-dependent glycerol-3-phosphate dehydrogenase signature 2.

CONSENSUS: G-G-K-x(2)-[GSTE]-Y-R-x(2)-A.

NAME: Mannitol dehydrogenases signature.

CONSENSUS: [LIVMY]-x-[FS]-x(2)-[STAGCV]-x-V-D-R-[TV]-x-[PS].

NAME: Histidinol dehydrogenase signature.

CONSENSUS: I-D-x(2)-A-G-P-[ST]-E-[LIVS]-[LIVMA](3)-[AC]-x(3)-A-x(4)-[LIVM]-[AV]-

CONSENSUS: [SACL]-[DE]-[LIVMFC]-[LIVM]-[SA]-x(2)-E-H.

NAME: L-lactate dehydrogenase active site.

CONSENSUS: [LIVMA]-G-[EQ]-H-G-[DN]-[ST].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases NAD-binding signature.

CONSENSUS: [LIVMA]-[AG]-[IVT]-[LIVMFY]-[AG]-x-G-[NHKRQGSAC]-[LIV]-G-x(13,14)-

CONSENSUS: [LIVMT]-x(2)-[FYwCTH]-[DNSTK].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases signature 2.

CONSENSUS: [LIVMFYWA]-[LIVFYWC]-x(2)-[SAC]-[DNQHR]-[IVFA]-[LIVF]-x-[LIVF]-[HNT]-x-

CONSENSUS: P-x(4)-[STN]-x(2)-[LIVMF]-x-[GSDN].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases signature 3.

CONSENSUS: [LMFATC]-[KPQ]-x-[GSTDN]-x-[LIVMFYWR]-[LIVMFYW](2)-N-x-[STAGC]-R-[GP]-x-

CONSENSUS: [LIVH]-[LIVMC]-[DNV].

NAME: 3-hydroxyisobutyrate dehydrogenase signature.

CONSENSUS: [LIVMFY](2)-G-L-G-x-[MQ]-G-x-[PGS]-[MA]-[SA].

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 1.

CONSENSUS: [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA].

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 2.

CONSENSUS: [LIVM]-G-x-[LIVM]-G-G-[AG]-T.

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 3.

CONSENSUS: A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x-[FYLH].

NAME: Hydroxymethylglutaryl-coenzyme A reductases profile.

NAME: 3-hydroxyacyl-CoA dehydrogenase signature.

CONSENSUS: [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-[PA]-[LIVMFY](2)-x(5)-

CONSENSUS: [LIVMFYCT]-[LIVMFY]-x(2)-[GV].

NAME: Malate dehydrogenase active site signature.

CONSENSUS: [LIVM]-T-[TRKMN]-L-D-x(2)-R-[STA]-x(3)-[LIVMFY].

NAME: Malic enzymes signature.

CONSENSUS: F-x-[DV]-D-x(2)-G-T-[GSA]-x-[IV]-x-[LIVMA]-[GAST](2)-[LIVMF](2).

NAME: Isocitrate and isopropylmalate dehydrogenases signature.

CONSENSUS: [NS]-[LIMYT]-[FYDN]-G-[DNT]-[IMVY]-x-[STGDN]-[DN]-x(2)-[SGAP]-x(3,4)-G-

CONSENSUS: [STG]-[LIVMPA]-G-[LIVMF].

NAME: 6-phosphogluconate dehydrogenase signature.

CONSENSUS: [LIVM]-x-D-x(2)-[GA]-[NQS]-K-G-T-G-x-W.

NAME: Glucose-6-phosphate dehydrogenase active site.

CONSENSUS: D-H-Y-L-G-K-[EQK].

NAME: IMP dehydrogenase / GMP reductase signature.

CONSENSUS: [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T.

NAME: Bacterial quinoprotein dehydrogenases signature 1.

CONSENSUS: [DEN]-W-x(3)-G-[RK]-x(6)-[FYW]-S-x(4)-[LIVM]-N-x(2)-N-V-x(2)-L-[RK].

NAME: Bacterial quinoprotein dehydrogenases signature 2.

CONSENSUS: W-x(4)-Y-D-x(3)-[DN]-[LIVMFY](4)-x(2)-G-x(2)-[STA]-P.

NAME: FMN-dependent alpha-hydroxy acid dehydrogenases active site.

CONSENSUS: S-N-H-G-[AG]-R-Q.

NAME: GMC oxidoreductases signature 1.

CONSENSUS: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]-x(2)-[PAG]-x(5)-

CONSENSUS: [DNESH].

NAME: GMC oxidoreductases signature 2.

CONSENSUS: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G.

NAME: Eukaryotic molybdopterin oxidoreductases signature.

CONSENSUS: [GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-

CONSENSUS: x(2)-[DE].

NAME: Prokaryotic molybdopterin oxidoreductases signature 1.

CONSENSUS: [STAN]-x-[CH]-x(2,3)-C-[STAG]-[GSTVMF]-x-C-x-[LIVMFYW]-x-[LIVMA]-x(3,4)-

CONSENSUS: [DENQKHT].

NAME: Prokaryotic molybdopterin oxidoreductases signature 2.

CONSENSUS: [STA]-x-[STAC](2)-x(2)-[STA]-D-[LIVMY](2)-L-P-x-[STAC](2)-x(2)-E.

NAME: Prokaryotic molybdopterin oxidoreductases signature 3.

CONSENSUS: A-x(3)-[GDT]-I-x-[DNQTK]-x-[DEA]-x-[LIVM]-x-[LIVMC]-x-[NS]-x(2)-[GS]-

CONSENSUS: x(5)-A-x-[LIVM]-[ST].

NAME: Aldehyde dehydrogenases glutamic acid active site.
 CONSENSUS: [LIVMFGA]-E-[LIMSTAC]-[GS]-G-[KNLM]-[SADN]-[TAPFV].

NAME: Aldehyde dehydrogenases cysteine active site.
 CONSENSUS: [FYLVA]-x(3)-G-[QE]-x-C-[LIVMGSTANC]-[AGCN]-x-[GSTADNEKR].

NAME: Aspartate-semialdehyde dehydrogenase signature.
 CONSENSUS: [LIVM]-[SADN]-x(2)-C-x-R-[LIVM]-x(4)-[GSC]-H-[STA].

NAME: Glyceraldehyde 3-phosphate dehydrogenase active site.
 CONSENSUS: [ASV]-S-C-[NT]-T-x(2)-[LIM].

NAME: N-acetyl-gamma-glutamyl-phosphate reductase active site.
 CONSENSUS: [LIVM]-[GSA]-x-P-G-C-[FY]-[AVP]-T-[GA]-x(3)-[GTAC]-[LIVM]-x-P.

NAME: Gamma-glutamyl phosphate reductase signature.
 CONSENSUS: V-x(5)-A-[LIV]-x-H-I-x(2)-[HY]-[GS]-[ST]-x-H-[ST]-[DE]-x-I.

NAME: Dihydrodipicolinate reductase signature.
 CONSENSUS: E-[IV]-x-E-x-H-x(3)-K-x-D-x-P-S-G-T-A.

NAME: Dihydroorotate dehydrogenase signature 1.
 CONSENSUS: [GS]-x(4)-[GK]-[STA]-[IVSTA]-[GT]-x(3)-[NQR]-x-G-[NH]-x(2)-P-[RT].

NAME: Dihydroorotate dehydrogenase signature 2.
 CONSENSUS: [LIV](2)-[GSA]-x-G-G-[IV]-x-[STGN]-x(3)-[ACV]-x(6)-G-A.

NAME: Coproporphyrinogen III oxidase signature.
 CONSENSUS: K-x-W-C-x(2)-[FYH](3)-[LIVM]-x-H-R-x-E-x-R-G-[LIVM]-G-G-[LIVM]-F-F-D.

NAME: Fumarate reductase / succinate dehydrogenase FAD-binding site.
 CONSENSUS: R-[ST]-H-[ST]-x(2)-A-x-G-G.

NAME: Acyl-CoA dehydrogenases signature 1.
 CONSENSUS: [GAC]-[LIVM]-[ST]-E-x(2)-[GSAN]-G-[ST]-D-x(2)-[GSA].

NAME: Acyl-CoA dehydrogenases signature 2.
 CONSENSUS: [QDE]-x(2)-G-[GS]-x-G-[LIVMFY]-x(2)-[DEN]-x(4)-[KR]-x(3)-[DEN].

NAME: Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 1.
 CONSENSUS: G-[LIVM]-P-x-E-x(3)-N-E-x(1,3)-R-V-A-x-[ST]-P-x-[GST]-V-x(2)-L-x-[KRH]-x-G.
 CONSENSUS: x-G.

NAME: Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 2.
 CONSENSUS: [LIVM](2)-G-[GA]-G-x-A-G-x(2)-[SA]-x(3)-[GA]-x-[SG]-[LIVM]-G-A-x-V-x(3)-D.
 CONSENSUS: x(3)-D.

NAME: Glu / Leu / Phe / Val dehydrogenases active site.
 CONSENSUS: [LIV]-x(2)-G-G-[SAG]-K-x-[GV]-x(3)-[DNST]-[PL].

NAME: D-amino acid oxidases signature.
 CONSENSUS: [LIVM](2)-H-[NHA]-Y-G-x-[GSA](2)-x-G-x(5)-G-x-A.

NAME: Pyridoxamine 5'-phosphate oxidase signature.
 CONSENSUS: [LIVF]-E-F-W-[QHG]-x(4)-R-[LIVM]-H-[DNE]-R.

NAME: Copper amine oxidase topaquinone signature.
 CONSENSUS: [LIVM]-[LIVMA]-[LIVM]-x(4)-T-x(2)-N-Y-[DE]-[YN].

NAME: Copper amine oxidase copper-binding site signature.
 CONSENSUS: T-x-G-x(2)-H-[LIVMF]-x(3)-E-[DE]-x-P.

NAME: Lysyl oxidase putative copper-binding region signature.
 CONSENSUS: W-E-W-H-S-C-H-Q-H-Y-H.

NAME: Delta 1-pyrroline-5-carboxylate reductase signature.
 CONSENSUS: [PALF]-x(2,3)-[LIV]-x(3)-[LIVM]-[STAC]-[STV]-x-[GAN]-G-x-T-x(2)-[AG]-[LIV]-x(2)-[LMF]-[DENQK].
 CONSENSUS: [LIV]-x(2)-[LMF]-[DENQK].

NAME: Dihydrofolate reductase signature.
 CONSENSUS: [LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYIV]-x(3)-[STIQ].

NAME: Tetrahydrofolate dehydrogenase/cyclohydrolase signature 1.
 CONSENSUS: [EQ]-x-[EQK]-[LIVM](2)-x(2)-[LIVM]-x(2)-[LIVMY]-N-x-[DN]-x(5)-[LIVMF](3)-

CONSENSUS: Q-L-P-[LV].

NAME: Tetrahydrofolate dehydrogenase/cyclohydrolase signature 2.

CONSENSUS: P-G-G-V-G-P-[MF]-T-[IV].

NAME: Oxygen oxidoreductases covalent FAD-binding site.

CONSENSUS: P-x(10)-[DE]-[LIVM]-x(3)-[LIVM]-x(9)-[LIVM]-x(3)-[GSA]-[GST]-G-H.

NAME: Pyridine nucleotide-disulphide oxidoreductases class-I active site.

CONSENSUS: G-G-x-C-[LIVA]-x(2)-G-C-[LIVM]-P.

NAME: Pyridine nucleotide-disulphide oxidoreductases class-II active site.

CONSENSUS: C-x(2)-C-D-[GA]-x(2,4)-[FY]-x(4)-[LIVM]-x-[LIVM](2)-G(3)-[DN].

NAME: Respiratory-chain NADH dehydrogenase subunit I signature 1.

CONSENSUS: G-[LIVMFYKRS]-[LIVMAGP]-Q-x-[LIVMFY]-x-D-[AGIM]-[LIVMFTA]-K-[LVMYST]-
[LIVMFYF]-x-[KR]-[EQG].

NAME: Respiratory-chain NADH dehydrogenase subunit I signature 2.

CONSENSUS: P-F-D-[LIVMFYQ]-[STAGPVM]-E-[GAC]-E-x-[EQ]-[LIVMS]-x(2)-G.

NAME: Respiratory-chain NADH dehydrogenase 20 Kd subunit signature.

CONSENSUS: [GN]-x-D-[KRST]-[LIVMF](2)-P-[IV]-D-[LIVMFYW](2)-x-P-x-C-P-[PT].

NAME: Respiratory-chain NADH dehydrogenase 24 Kd subunit signature.

CONSENSUS: D-x(2)-F-[ST]-x(5)-C-L-G-x-C-x(2)-[GA]-P.

NAME: Respiratory chain NADH dehydrogenase 30 Kd subunit signature.

CONSENSUS: E-R-E-x(2)-[DE]-[LIVMF](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMS].

NAME: Respiratory chain NADH dehydrogenase 49 Kd subunit signature.

CONSENSUS: [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMT]-x-E-x-[KRQ].

NAME: Respiratory-chain NADH dehydrogenase 51 Kd subunit signature 1.

CONSENSUS: G-[AM]-G-[AR]-Y-[LIVM]-C-G-[DE](2)-[STA](2)-[LIM](2)-[EN]-S.

NAME: Respiratory-chain NADH dehydrogenase 51 Kd subunit signature 2.

CONSENSUS: E-S-C-G-x-C-x-P-C-R-x-G.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 1.

CONSENSUS: P-x(2)-C-[YWS]-x(7)-G-x-C-R-x-C.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 2.

CONSENSUS: C-P-x-C-[DE]-x-[GS](2)-x-C-x-L-Q.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 3.

CONSENSUS: R-C-[LIVM]-x-C-x-R-C-[LIVM]-x-[FY].

NAME: Nitrite and sulfite reductases iron-sulfur/siroheme-binding site.

CONSENSUS: [STV]-G-C-x(3)-C-x(6)-[DE]-[LIVMF]-[GAT]-[LIVMF].

NAME: Uricase signature.

CONSENSUS: L-x-[LV]-L-K-[ST]-T-x-S-x-F-x(2)-[FY]-x(4)-[FY].

NAME: Heme-copper oxidase catalytic subunit, copper B binding region signature.

CONSENSUS: [YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H.

NAME: CO II and nitrous oxide reductase dinuclear copper centers signature.

CONSENSUS: V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M.

NAME: Cytochrome c oxidase subunit Vb, zinc binding region signature.

CONSENSUS: [LIVM](2)-[FYW]-x(10)-C-x(2)-C-G-x(2)-[FY]-K-L.

NAME: Multicopper oxidases signature 1.

CONSENSUS: G-x-[FYW]-x-[LIVMFYW]-x-[CST]-x(8)-G-[LM]-x(3)-[LIVMFYW].

NAME: Multicopper oxidases signature 2.

CONSENSUS: H-C-H-x(3)-H-x(3)-[AG]-[LM].

NAME: Peroxidases proximal heme-ligand signature.

CONSENSUS: [DET]-[LIVMTA]-x(2)-[LIVM]-[LIVMSTAG]-[SAG]-[LIVMSTAG]-H-[STA]-[LIVMFY].

NAME: Peroxidases active site signature.

CONSENSUS: [SGATV]-x(3)-[LIVMA]-R-[LIVMA]-x-[FW]-H-x-[SAC].

NAME: Catalase proximal heme-ligand signature.
 CONSENSUS: R-[LIVMFSTAN]-F-[GASTNP]-Y-x-D-[AST]-[QEH].

NAME: Catalase proximal active site signature.
 CONSENSUS: [IF]-x-[RH]-x(4)-[EQ]-R-x(2)-H-x(2)-[GAS]-[GASTF]-[GAST].

NAME: Glutathione peroxidases selenocysteine active site.
 CONSENSUS: [GN]-[RKHNFCY]-x-[LIVMFC]-[LIVMF](2)-x-N-[VT]-x-[STC]-x-C-[GA]-x-T.

NAME: Glutathione peroxidases signature 2.
 CONSENSUS: [LIV]-[AGD]-F-P-[CS]-[NG]-Q-F.

NAME: Lipoxygenases iron-binding region signature 1.
 CONSENSUS: H-[EQ]-x(3)-H-x-[LM]-[NQRC]-[GST]-H-[LIVMSTAC](3)-E.

NAME: Lipoxygenases iron-binding region signature 2.
 CONSENSUS: [LIVMA]-H-P-[LIVM]-x-[KRQ]-[LIVMF](2)-x-[AP]-H.

NAME: Extradial ring-cleavage dioxygenases signature.
 CONSENSUS: [GNTIV]-x-H-x(5,7)-[LIVMF]-Y-x(2)-[DENTA]-P-x-[GP]-x(2,3)-E.

NAME: Intradiol ring-cleavage dioxygenases signature.
 CONSENSUS: [LIVM]-x-G-x-[LIVM]-x(4)-[GS]-x(2)-[LIVM]-x(4)-[LIVM]-[DE]-[LIVMFY]-
 CONSENSUS: x(6)-G-x-[FY].

NAME: Indoleamine 2,3-dioxygenase signature 1.
 CONSENSUS: G-G-S-[AN]-[GA]-Q-S-S-x(2)-Q.

NAME: Indoleamine 2,3-dioxygenase signature 2.
 CONSENSUS: [FY]-L-[DQ]-[DE]-[LIVM]-x(2)-Y-M-x(3)-H-[KR].

NAME: Bacterial ring hydroxylating dioxygenases alpha-subunit signature.
 CONSENSUS: C-x-H-R-[GA]-x(8)-G-N-x(5)-C-x-[FY]-H.

NAME: Bacterial luciferase subunits signature.
 CONSENSUS: [GA]-[LIVM]-P-[LIVM]-x-[LIVMFY]-x-W-x(6)-[RK]-x(6)-Y-x(3)-[AR].

NAME: ubiH/COQ6 monooxygenase family signature.
 CONSENSUS: H-P-[LIV]-[AG]-G-Q-G-x-N-x-G-x(2)-D.

NAME: Bioprotein-dependent aromatic amino acid hydroxylases signature.
 CONSENSUS: P-D-x(2)-H-[DE]-[LI]-[LIVMF]-G-H-[LIVMC]-P.

NAME: Copper type II, ascorbate-dependent monooxygenases signature 1.
 CONSENSUS: H-H-M-x(2)-F-x-C.

NAME: Copper type II, ascorbate-dependent monooxygenases signature 2.
 CONSENSUS: H-x-F-x(4)-H-T-H-x(2)-G.

NAME: Tyrosinase CuA-binding region signature.
 CONSENSUS: H-x(4,5)-F-[LIVMFTP]-x-[FW]-H-R-x(2)-[LM]-x(3)-E.

NAME: Tyrosinase and hemocyanins CuB-binding region signature.
 CONSENSUS: D-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D.

NAME: Fatty acid desaturases family 1 signature.
 CONSENSUS: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y.

NAME: Fatty acid desaturases family 2 signature.
 CONSENSUS: [ST]-[SA]-x(3)-[QR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]-[DE].

NAME: Cytochrome P450 cysteine heme-iron ligand signature.
 CONSENSUS: [FW]-[SGNH]-x-[GD]-x-[RHPT]-x-C-[LIVMFAP]-[GAD].

NAME: Heme oxygenase signature.
 CONSENSUS: L-L-V-A-H-A-Y-T-R.

NAME: Copper/Zinc superoxide dismutase signature 1.
 CONSENSUS: [GA]-[IFAT]-H-[LIVF]-H-x(2)-[GP]-[SDG]-x-[STAGD].

NAME: Copper/Zinc superoxide dismutase signature 2.
 CONSENSUS: G-[GN]-[SGA]-G-x-R-x-[SGA]-C-x(2)-[IV].

NAME: Manganese and iron superoxide dismutases signature.

CONSENSUS: D-x-W-E-H-[STA]-[FY](2).

NAME: Ribonucleotide reductase large subunit signature.

CONSENSUS: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-[PA].

NAME: Ribonucleotide reductase small subunit signature.

CONSENSUS: [IVMSEQ]-E-x(1,2)-[LIVTA]-[HY]-[GSA]-x-[STAVM]-Y-x(2)-[LIVMQ]-x(3)-[LIFY]-[IVFYCSA].

NAME: Nitrogenases component I alpha and beta subunits signature 1.

CONSENSUS: [LIVMFYH]-[LIVMFST]-H-[AG]-[AGSP]-[LIVMNQA]-[AG]-C.

NAME: Nitrogenases component I alpha and beta subunits signature 2.

CONSENSUS: [STANQ]-[ET]-C-x(5)-G-D-[DN]-[LIVMT]-x-[STAGR]-[LIVMFYST].

NAME: NifH/frxC family signature 1.

CONSENSUS: E-x-G-G-P-x(2)-[GA]-x-G-C-[AG]-G.

NAME: NifH/frxC family signature 2.

CONSENSUS: D-x-L-G-D-V-V-C-G-G-F-[AG]-x-P.

NAME: Nickel-dependent hydrogenases large subunit signature 1.

CONSENSUS: R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C.

NAME: Nickel-dependent hydrogenases large subunit signature 2.

CONSENSUS: [FY]-D-P-C-[LIM]-[ASG]-C-x(2,3)-H.

NAME: Glutamyl-tRNA reductase signature.

CONSENSUS: H-[LIVM]-x(2)-[LIVM]-[GSTAC](3)-[LIVM]-[DEQ]-S-[LIVMA]-[LIVM](2)-[GF]-E-x-[QR]-[IV]-[LIT]-[STAG]-Q-[LIVM]-[KR].

NAME: Bacterial-type phytoene dehydrogenase signature.

CONSENSUS: [NG]-x-[FYWV]-[LIVMF]-x-G-[AGC]-[GS]-[TA]-[HQT]-P-G-[STAV]-G-[LIVM]-x(5)-[GS].

NAME: Glycine radical signature.

CONSENSUS: [STIV]-x-R-[IVT]-[CSA]-G-Y-x-[GACV].

NAME: Ergosterol biosynthesis ERG4/ERG24 family signature 1.

CONSENSUS: G-x(2)-[LIVM]-Y-D-x-[FY]-x-G-x(2)-L-N-P-R.

NAME: Ergosterol biosynthesis ERG4/ERG24 family signature 2.

CONSENSUS: [LIVM](2)-H-R-x(2)-R-D-x(3)-C-x(2)-K-Y-G.

NAME: NNMT/PNMT/TEMT family of methyltransferases signature.

CONSENSUS: L-I-D-I-G-S-G-P-T-[IV]-Y-Q-L-L-S-A-C.

NAME: RNA methyltransferase trmA family signature 1.

CONSENSUS: [DN]-P-[PA]-R-x-G-x(14,16)-[LIVM](2)-Y-x-S-C-N-x(2)-T.

NAME: RNA methyltransferase trmA family signature 2.

CONSENSUS: [LIVMF]-D-x-F-P-[QHY]-[ST]-x-H-[LIVMFY]-E.

NAME: Thymidylate synthase active site.

CONSENSUS: R-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]-x(3)-[QMT]-[FYW]-x-[LV].

NAME: Ribosomal RNA adenine dimethylases signature.

CONSENSUS: [LIVM]-[LIVMFY]-[DE]-x-G-[STAPV]-G-x-[GA]-x-[LIVMF]-[ST]-x(2)-[LIVM]-x(6)-[LIVMY]-x-[STAGV]-[LIVMFYHC]-E-x-D.

NAME: Methylated-DNA-protein-cysteine methyltransferase active site.

CONSENSUS: [LIVMF]-P-C-H-R-[LIVMF](2).

NAME: N-6 Adenine-specific DNA methylases signature.

CONSENSUS: [LIVMAC]-[LIVFYWA]-x-[DN]-P-P-[FYW].

NAME: N-4 cytosine-specific DNA methylases signature.

CONSENSUS: [LIVMF]-T-S-P-P-[FY].

NAME: C-5 cytosine-specific DNA methylases active site.

CONSENSUS: [DENKS]-x-[FLIV]-x(2)-[GSTC]-x-P-C-x(2)-[FYWLIM]-S.

NAME: C-5 cytosine-specific DNA methylases C-terminal signature.

CONSENSUS: [RKQGT]-x(2)-G-N-[STAG]-[LIVMF]-x(3)-[LIVMT]-x(3)-[LIVM]-x(3)-[LIVM].

NAME: Protein-L-isoaspartate(D-aspartate) O-methyltransferase signature.

CONSENSUS: [GSA]-D-G-x(2)-G-[FYWV]-x(3)-[AS]-P-[FY]-[DN]-x-I.

NAME: Uroporphyrin-III C-methyltransferase signature 1.

CONSENSUS: [LIVM]-[GS]-[STAL]-G-P-G-x(3)-[LIVMFY]-[LIVM]-T-[LIVM]-[KRHQ]-[AG].

NAME: Uroporphyrin-III C-methyltransferase signature 2.

CONSENSUS: V-x(2)-[LI]-x(2)-G-D-x(3)-[FYW]-[GS]-x(8)-[LIVF]-x(5,6)-[LIVMFYWPAC]-
x-[LIVMY]-x-P-G.

NAME: ubiE/COQ5 methyltransferase family signature 1.

CONSENSUS: Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W.

NAME: ubiE/COQ5 methyltransferase family signature 2.

CONSENSUS: R-V-[LIVM]-K-[PV]-G-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S.

NAME: Serine hydroxymethyltransferase pyridoxal-phosphate attachment site.

CONSENSUS: [DEH]-[LIVMFY]-x-[STMV]-[GST]-[ST](2)-H-K-[ST]-[LF]-x-G-[PAC]-[RQ]-
[GSA]-[GA].

NAME: Phosphoribosylglycinamide formyltransferase active site.

CONSENSUS: G-x-[STM]-[IVT]-x-[FYWVQ]-[VMAT]-x-[DEV]-x-[LIVMY]-D-x-G-x(2)-[LIVT]-
x(6)-[LIVM].

NAME: Aspartate and ornithine carbamoyltransferases signature.

CONSENSUS: F-x-[EK]-x-S-[GT]-R-T.

NAME: Transketolase signature 1.

CONSENSUS: R-x(3)-[LIVMTA]-[DENQSTHKF]-x(5,6)-[GSN]-G-H-[PLIVMF]-[GSTA]-x(2)-
[LMC]-[GS].

NAME: Transketolase signature 2.

CONSENSUS: G-[DEQSA]-[DN]-G-[PAEQ]-[ST]-[HQ]-x-[PAGM]-[LIVMYAC]-[DEFYW]-x(2)-
[STAP]-x(2)-[RGA].

NAME: Transaldolase signature 1.

CONSENSUS: [DG]-[IVSA]-T-[ST]-N-P-[STA]-[LIVMF](2).

NAME: Transaldolase active site.

CONSENSUS: [LIVM]-x-[LIVM]-K-[LIVM]-[PAS]-x-[ST]-x-[DENQPAS]-G-[LIVM]-x-[AGV]-x-
[QEKST]-x-[LIVM].

NAME: Acyltransferases ChoActase / COT / CPT family signature 1.

CONSENSUS: [LI]-P-x-[LVP]-P-[IVTA]-P-x-[LIVM]-x-[DENQAS]-[ST]-[LIVM]-x(2)-[LY].

NAME: Acyltransferases ChoActase / COT / CPT family signature 2.

CONSENSUS: R-[FYW]-x-[DA]-[KA]-x(0,1)-[LIVMFY]-x-[LIVMFY](2)-x(3)-[DNS]-[GSA]-x(6)-
[DE]-[HS]-x(3)-[DE]-[GA].

NAME: Thiolases acyl-enzyme intermediate signature.

CONSENSUS: [LIVM]-[NST]-x(2)-C-[SAGLI]-[ST]-[SAG]-[LIVMFYNS]-x-[STAG]-[LIVM]-x(6)-
[LIVM].

NAME: Thiolases signature 2.

CONSENSUS: N-x(2)-G-G-x-[LIVM]-[SA]-x-G-H-P-x-G-x-[ST]-G.

NAME: Thiolases active site.

CONSENSUS: [AG]-[LIVMA]-[STAGLIVM]-[STAG]-[LIVMA]-C-x-[AG]-x-[AG]-x-[AG]-x-[SAG].

NAME: Chloramphenicol acetyltransferase active site.

CONSENSUS: Q-[LIV]-H-H-[SA]-x(2)-D-G-[FY]-H.

NAME: Hexapeptide-repeat containing-transferases signature.

CONSENSUS: [LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIVAC]-x-[LIV]-[GAED]-x(2)-
[STAVR]-x-[LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIV].

NAME: Beta-ketoacyl synthases active site.

CONSENSUS: G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF].

NAME: Chalcone and stilbene synthases active site.

CONSENSUS: R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GA]-G-[GA]-[STAV]-x-[LIVMF]-
 CONSENSUS: [RA].

NAME: Myristoyl-CoA:protein N-myristoyltransferase signature 1.
 CONSENSUS: E-I-N-F-L-C-x-H-K.

NAME: Myristoyl-CoA:protein N-myristoyltransferase signature 2.
 CONSENSUS: K-F-G-x-G-D-G.

NAME: Gamma-glutamyltranspeptidase signature.
 CONSENSUS: T-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-
 CONSENSUS: x(1,2)-[FY]-G.

NAME: Transglutaminases active site.
 CONSENSUS: [GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G.

NAME: Phosphorylase pyridoxal-phosphate attachment site.
 CONSENSUS: E-A-[SC]-G-x-[GS]-x-M-K-x(2)-[LM]-N.

NAME: UDP-glycosyltransferases signature.
 CONSENSUS: {FW}-x(2)-Q-x(2)-[LIVMYA]-[LIMV]-x(4,6)-[LVGAC]-[LVFYA]-[LIVMF]-[STAGCM]-
 CONSENSUS: [HNQ]-[STAGC]-G-x(2)-[STAG]-x(3)-[STAGL]-[LIVMFA]-x(4)-[PQR]-[LIVMT]-
 CONSENSUS: x(3)-[PA]-x(3)-[DES]-[QEHN].

NAME: Purine/pyrimidine phosphoribosyl transferases signature.
 CONSENSUS: [LIVMFYWCTA]-[LIVM]-[LIVMA]-[LIVMFC]-[DE]-D-[LIVMS]-[LIVM]-[STAVD]-
 CONSENSUS: [STAR]-[GAC]-x-[STAR].

NAME: Glutamine amidotransferases class-I active site.
 CONSENSUS: {PAS}-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]-x-[LIVMFA].

NAME: Glutamine amidotransferases class-II active site.
 CONSENSUS: <x(0,11)-C-[GS]-[IV]-[LIVMFYW]-[AG].

NAME: Purine and other phosphorylases family 1 signature.
 CONSENSUS: [GST]-x-G-[LIVM]-G-x-[PA]-S-x-[GSTA]-I-x(3)-E-L.

NAME: Purine and other phosphorylases family 2 signature.
 CONSENSUS: [LIV]-x(3)-G-x(2)-H-x-[LIVMFY]-x(4)-[LIVMF]-x(3)-[ATV]-x(1,2)-[LIVM]-x-
 CONSENSUS: [ATV]-x(4)-[GN]-x(3,4)-[LIVMF](2)-x(2)-[STN]-[SA]-x-G-[GS]-[LIVM].

NAME: Thymidine and pyrimidine-nucleoside phosphorylases signature.
 CONSENSUS: S-[GS]-R-[GA]-[LIV]-x(2)-[TA]-[GA]-G-T-x-D-x-[LIV]-E.

NAME: ATP phosphoribosyltransferase signature.
 CONSENSUS: E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM].

NAME: NAD:arginine ADP-ribosyltransferases signature.
 CONSENSUS: [FY]-x-[FY]-K-x(2)-H-[FY]-x-L-[ST]-x-A.

NAME: Prolipoprotein diacylglycerol transferase signature.
 CONSENSUS: G-R-x-[GA]-N-F-[LIVMF]-N-x-E-x(2)-G.

NAME: S-adenosylmethionine synthetase signature 1.
 CONSENSUS: G-A-G-D-Q-G-x(3)-G-Y.

NAME: S-adenosylmethionine synthetase signature 2.
 CONSENSUS: G-[GA]-G-[ASC]-F-S-x-K-[DE].

NAME: Polyprenyl synthetases signature 1.
 CONSENSUS: [LIVM](2)-x-D-D-x(2,4)-D-x(4)-R-R-[GH].

NAME: Polyprenyl synthetases signature 2.
 CONSENSUS: [LIVMFY]-G-x(2)-[FYI]-Q-[LIVM]-x-D-D-[LIVMFY]-x-[DNG].

NAME: Squalene and phytoene synthases signature 1.
 CONSENSUS: Y-[CSAM]-x(2)-[VSG]-A-[GSA]-[LIVAT]-[IV]-G-x(2)-[LMSC]-x(2)-[LIV].

NAME: Squalene and phytoene synthases signature 2.
 CONSENSUS: [LIVM]-G-x(3)-Q-x(2,3)-N-[IF]-x-R-D-[LIVMFY]-x(2)-[DE]-x(4,7)-R-x-[FY]-
 CONSENSUS: x-P.

NAME: Protein prenyltransferases alpha subunit repeat signature.
 CONSENSUS: [PSIAV]-x-[NDFV]-[NEQIY]-x-[LIVMAGP]-W-[NQSTHF]-[FYHQ]-[LIVMR].

NAME: Riboflavin synthase alpha chain family signature.

CONSENSUS: [LIVMF]-x(5)-G-[STADNQ]-[KREQIYW]-V-N-[LIVM]-E.

NAME: Dihydropteroate synthase signature 1.

CONSENSUS: [LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-x-[SG].

NAME: Dihydropteroate synthase signature 2.

CONSENSUS: [GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)-[STA]-x-P.

NAME: EPSP synthase signature 1.

CONSENSUS: [LIVM]-x(2)-[GN]-N-[SA]-G-T-[STA]-x-R-x-[LIVMY]-x-[GSTA].

NAME: EPSP synthase signature 2.

CONSENSUS: [KR]-x-[KH]-E-[CST]-[DNE]-R-[LIVM]-x-[STA]-[LIVMC]-x(2)-[EN]-[LIVMF]-x-

CONSENSUS: [KRA]-[LIVMF]-G.

NAME: FLAP/GST2/LTC4S family signature.

CONSENSUS: G-x(3)-F-E-R-V-[FY]-x-A-[NQ]-x-N-C.

NAME: Aminotransferases class-I pyridoxal-phosphate attachment site.

CONSENSUS: [GS]-[LIVMFYTAC]-[GSTA]-K-x(2)-[GSALVN]-[LIVMFA]-x-[GNAR]-x-R-[LIVMA]-

CONSENSUS: [GA].

NAME: Aminotransferases class-II pyridoxal-phosphate attachment site.

CONSENSUS: T-[LIVMFYW]-[STAG]-K-[SAG]-[LIVMFYWR]-[SAG]-x(2)-[SAG].

NAME: Aminotransferases class-III pyridoxal-phosphate attachment site.

CONSENSUS: [LIVMFYWC](2)-x-D-E-[LIVMA]-x(2)-[GP]-x(0,1)-[LIVMFYWAG]-x(0,1)-[SACR]-x-

CONSENSUS: [GSAD]-x(12,16)-D-[LIVMFYWC]-x(2,3)-[GSA]-K-x(3)-[GSTADN]-[GSA].

NAME: Aminotransferases class-IV signature.

CONSENSUS: E-x-[STAGCI]-x(2)-N-[LIVMFAC]-[FY]-x(6,12)-[LIVMF]-x-T-x(6,8)-[LIVM]-x-

CONSENSUS: [GS]-[LIVM]-x-[KR].

NAME: Aminotransferases class-V pyridoxal-phosphate attachment site.

CONSENSUS: [LIVFYCHT]-[DGH]-[LIVMFYAC]-[LIVMFYA]-x(2)-[GSTAC]-[GSTA]-[HQR]-K-

CONSENSUS: x(4,6)-G-x-[GSAT]-x-[LIVMFYSAC].

NAME: Hexokinases signature.

CONSENSUS: [LIVM]-G-F-[TN]-F-S-[FY]-P-x(5)-[LIVM]-[DNST]-x(3)-[LIVM]-x(2)-W-T-K-x-

CONSENSUS: [LF].

NAME: Galactokinase signature.

CONSENSUS: G-R-x-N-[LIV]-I-G-E-H-x-D-Y.

NAME: GHMP kinases putative ATP-binding domain.

CONSENSUS: [LIVM]-[PK]-x-[GSTA]-x(0,1)-G-L-[GS]-S-S-[GSA]-[GSTAC].

NAME: Phosphofructokinase signature.

CONSENSUS: [RK]-x(4)-G-H-x-Q-[QR]-G-G-x(5)-D-R.

NAME: pfkB family of carbohydrate kinases signature 1.

CONSENSUS: [AG]-G-x(0,1)-[GAP]-x-N-x-[STA]-x(6)-[GS]-x(9)-G.

NAME: pfkB family of carbohydrate kinases signature 2.

CONSENSUS: [DNSK]-[PSTV]-x-[SAG](2)-[GD]-D-x(3)-[SAGV]-[AG]-[LIVMFY]-[LIVMSTAP].

NAME: ROK family signature.

CONSENSUS: [LIVM]-x(2)-G-[LIVMFCT]-G-x-[GA]-[LIVMFA]-x(8)-G-x(3,5)-[GATP]-x(2)-

CONSENSUS: G-[RKH].

NAME: Phosphoribulokinase signature.

CONSENSUS: K-[LIVM]-x-R-D-x(3)-R-G-x-[ST]-x-E.

NAME: Thymidine kinase cellular-type signature.

CONSENSUS: [GA]-x(1,2)-[DE]-x-Y-x-[STAP]-x-C-[NKR]-x-[CH]-[LIVMFYWH].

NAME: FGGY family of carbohydrate kinases signature 1.

CONSENSUS: [MFYGS]-x-[PST]-x(2)-K-[LIVMFYW]-x-W-[LIVMF]-x-[DENQTKR]-[ENQH].

NAME: FGGY family of carbohydrate kinases signature 2.

CONSENSUS: [GSA]-x-[LIVMFYW]-x-G-[LIVM]-x(7,8)-[HDENQ]-[LIVMF]-x(2)-[AS]-[STAIVM]-

CONSENSUS: [LIVMFY]-[DEQ].

NAME: Protein kinases ATP-binding region signature.

CONSENSUS: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-
CONSENSUS: x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K.

NAME: Serine/Threonine protein kinases active-site signature.

CONSENSUS: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3).

NAME: Tyrosine protein kinases specific active-site signature.

CONSENSUS: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3).

NAME: Protein kinase domain profile.

NAME: Casein kinase II regulatory subunit signature.

CONSENSUS: C-P-x-[LIVMY]-x-C-x(5)-L-P-[LIVMC]-G-x(9)-V-[KR]-x(2)-C-P-x-C.

NAME: Pyruvate kinase active site signature.

CONSENSUS: [LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQH]-[GSTA]-[LIVM].

NAME: Shikimate kinase signature.

CONSENSUS: [KR]-x(2)-E-x(3)-[LIVMF]-x(8,12)-[LIVMF](2)-[SA]-x-G(3)-x-[LIVMF].

NAME: Prokaryotic diacylglycerol kinase signature.

CONSENSUS: E-x-[LIVM]-N-[ST]-[SA]-[LIV]-E-x(2)-V-D.

NAME: Phosphatidylinositol 3- and 4-kinases signature 1.

CONSENSUS: [LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q.

NAME: Phosphatidylinositol 3- and 4-kinases signature 2.

CONSENSUS: [GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x-D-R-H-x(2)-N.

NAME: Acetate and butyrate kinases family signature 1.

CONSENSUS: [LIVM](2)-x-[LIVM]-N-x-G-S-[ST]-S-x-[KE].

NAME: Acetate and butyrate kinases family signature 2.

CONSENSUS: [LIVMA](2)-x(2)-H-x-G-x-G-x-[ST]-[LIVM]-x-[AV]-x(3)-G.

NAME: Phosphoglycerate kinase signature.

CONSENSUS: [KRHGTCTV]-[VT]-[LIVMF]-[LIVMC]-R-x-D-x-N-[SACV]-P.

NAME: Aspartokinase signature.

CONSENSUS: [LIVM]-x-K-[FY]-G-G-[ST]-[SC]-[LIVM].

NAME: Glutamate 5-kinase signature.

CONSENSUS: [GSTN]-x(2)-G-x-G-[GC]-[IM]-x-[STA]-K-[LIVM]-x-[SA]-[TCA]-x(2)-[GALV]-
CONSENSUS: x(3)-G.

NAME: ATP:guanido phosphotransferases active site.

CONSENSUS: C-P-x(0,1)-[ST]-N-[IL]-G-T.

NAME: PTS HPR component histidine phosphorylation site signature.

CONSENSUS: G-[LIVM]-H-[STA]-R-[PA]-[GSTA]-[STAM].

NAME: PTS HPR component serine phosphorylation site signature.

CONSENSUS: [GSADE]-[KREQTV]-x(4)-[KRN]-S-[LIVMF](2)-x-[LIVM]-x(2)-[LIVM]-[GAD].

NAME: PTS EIIA domains phosphorylation site signature 1.

CONSENSUS: G-x(2)-[LIVMF](3)-H-[LIVMF]-G-[LIVMF]-x-T-[ALV].

NAME: PTS EIIA domains phosphorylation site signature 2.

CONSENSUS: [DENQ]-x(6)-[LIVMF]-[GA]-x(2)-[LIVM]-A-[LIVM]-P-H-[GAC].

NAME: PTS EIIB domains cysteine phosphorylation site signature.

CONSENSUS: N-[LIVMFY]-x(5)-C-x-T-R-[LIVMF]-x-[LIVMF]-x-[LIVM]-x-[DQ].

NAME: Adenylate kinase signature.

CONSENSUS: [LIVMFYW](3)-D-G-[FYI]-P-R-x(3)-[NQ].

NAME: Nucleoside diphosphate kinases active site.

CONSENSUS: N-x(2)-H-[GA]-S-D-[SA]-[LIVMPKNE].

NAME: Guanylate kinase signature.

CONSENSUS: T-[ST]-R-x(2)-[KR]-x(2)-[DE]-x(2)-G-x(2)-Y-x-[FY]-[LIVMK].

NAME: Guanylate kinase domain profile.

NAME: Phosphoribosyl pyrophosphate synthetase signature.

CONSENSUS: D-[LI]-H-[SA]-x-Q-[IMST]-[QM]-G-[FY]-F-x(2)-P-[LIVMFC]-D.

NAME: 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature.

CONSENSUS: G-[PE]-R-x(2)-D-L-D-[LIVM](2).

NAME: Bacteriophage-type RNA polymerase family active site signature 1.

CONSENSUS: P-[LIVM]-x(2)-D-[GA]-[ST]-[AC]-[SN]-[GA]-[LIVMFY]-Q.

NAME: Bacteriophage-type RNA polymerase family active site signature 2.

CONSENSUS: [LIVMF]-x-R-x(3)-K-x(2)-[LIVMF]-M-[PT]-x(2)-Y.

NAME: Eukaryotic RNA polymerase II heptapeptide repeat.

CONSENSUS: Y-[ST]-P-[ST]-S-P-[STANK].

NAME: RNA polymerases beta chain signature.

CONSENSUS: G-x-K-[LIVMFA]-[STAC]-[GSTN]-x-[HSTA]-[GS]-[QNH]-K-G-[IVT].

NAME: RNA polymerases M / 15 Kd subunits signature.

CONSENSUS: F-C-x-[DEKST]-C-[GNK]-[DNSA]-[LIVMH]-[LIVM]-x(8,14)-C-x(2)-C.

NAME: RNA polymerases D / 30 to 40 Kd subunits signature.

CONSENSUS: N-[SGA]-[LIVMF]-R-R-x(9)-[SA]-x(3)-V-x(4)-N-x-[STA]-x(3)-[DN]-E-x-[LI]-

CONSENSUS: [GA]-x-R-[LI]-[GA]-[LIVM](2)-P.

NAME: RNA polymerases H / 23 Kd subunits signature.

CONSENSUS: H-[NEI]-[LIVM]-V-P-x-H-x(2)-[LIVM]-x(2)-[DE].

NAME: RNA polymerases K / 14 to 18 Kd subunits signature.

CONSENSUS: [ST]-x-[FY]-E-x-[AT]-R-x-[LIVM]-[GSA]-x-R-[SA]-x-Q.

NAME: RNA polymerases L / 13 to 16 Kd subunits signature.

CONSENSUS: [DE](2)-H-[ST]-[LIVM]-[GAP]-N-x(11)-V-x-[FM]-x(2)-Y-x(3)-H-P.

NAME: RNA polymerases N / 8 Kd subunits signature.

CONSENSUS: [LIVMF](2)-P-[LIVM]-x-C-F-[ST]-C-G.

NAME: DNA polymerase family A signature.

CONSENSUS: R-x(2)-[GSAV]-K-x(3)-[LIVMFY]-[AGQ]-x(2)-Y-x(2)-[GS]-x(3)-[LIVMA].

NAME: DNA polymerase family B signature.

CONSENSUS: [YA]-[GLIVMSTAC]-D-T-D-[SG]-[LIVMFTC]-x-[LIVMSTAC].

NAME: DNA polymerase family X signature.

CONSENSUS: G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D-[LIVMFY](3)-x(2)-[SAP].

NAME: Galactose-1-phosphate uridyl transferase family 1 active site signature.

CONSENSUS: F-E-N-[RK]-G-x(3)-G-x(4)-H-P-H-x-Q.

NAME: Galactose-1-phosphate uridyl transferase family 2 signature.

CONSENSUS: D-L-P-I-V-G-G-[ST]-[LIVM](2)-[SA]-H-[DEN]-H-[FY]-Q-G-G.

NAME: ADP-glucose pyrophosphorylase signature 1.

CONSENSUS: [AG]-G-G-x-G-[STK]-x-L-x(2)-L-[TA]-x(3)-A-x-P-A-[LV].

NAME: ADP-glucose pyrophosphorylase signature 2.

CONSENSUS: W-[FY]-x-G-[ST]-A-[DNSH]-[AS]-[LIVMFYW].

NAME: ADP-glucose pyrophosphorylase signature 3.

CONSENSUS: [APV]-[GS]-M-G-[LIVMN]-Y-[IVC]-[LIVMFY]-x(2)-[DENPHK].

NAME: Phosphatidate cytidyltransferase signature.

CONSENSUS: S-x-[LIVMF]-K-R-x(4)-K-D-x-[GSA]-x(2)-[LI]-[PG]-x-H-G-G-[LIVM]-x-D-R-

CONSENSUS: [LIVMFT]-D.

NAME: Ribonuclease PH signature.

CONSENSUS: C-[DE]-[LIVM](2)-Q-[GTA]-D-G-[SG]-x(2)-[TA]-A.

NAME: 2'-5'-oligoadenylate synthetases signature 1.

CONSENSUS: G-G-S-x-[AG]-[KR]-x-T-x-L-[KR]-[GST]-x-S-D-[AG].

NAME: 2'-5'-oligoadenylate synthetases signature 2.

CONSENSUS: R-P-V-I-L-D-P-x-[DE]-P-T.

NAME: CDP-alcohol phosphatidyltransferases signature.

CONSENSUS: D-G-x(2)-A-R-x(8)-G-x(3)-D-x(3)-D.

NAME: PEP-utilizing enzymes phosphorylation site signature.

CONSENSUS: G-[GA]-x-[TN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG].

NAME: PEP-utilizing enzymes signature 2.

CONSENSUS: [DEQS]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q-[LIVMFYGT]-[STALIV]-

CONSENSUS: [LIVMF]-[GAS]-x(2)-R.

NAME: Rhodanese signature 1.

CONSENSUS: [FY]-x(3)-H-[LIV]-P-G-A-x(2)-[LIVF].

NAME: Rhodanese C-terminal signature.

CONSENSUS: [AV]-x(2)-[FY]-[DEAP]-G-[GSA]-[WF]-x-E-[FYW].

NAME: CoA transferases signature 1.

CONSENSUS: [DN]-[GN]-x(2)-[LIVMFA](3)-G-G-F-x(3)-G-x-P.

NAME: CoA transferases signature 2.

CONSENSUS: [LF]-[HQ]-S-E-N-G-[LIVF](2)-[GA].

NAME: Phospholipase A2 histidine active site.

CONSENSUS: C-C-x(2)-H-x(2)-C.

NAME: Phospholipase A2 aspartic acid active site.

CONSENSUS: [LIVMA]-C-[LIVMFYWPCST]-C-D-x(5)-C.

NAME: Lipases, serine active site.

CONSENSUS: [LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-S-x-G-[GSTAC].

NAME: Colipase signature.

CONSENSUS: Y-x(2)-Y-Y-x-C-x-C.

NAME: Lipolytic enzymes "G-D-S-L" family, serine active site.

CONSENSUS: [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G.

NAME: Lipolytic enzymes "G-D-X-G" family, putative histidine active site.

CONSENSUS: [LIVMF](2)-x-[LIVMF]-H-G-G-[SAG]-[FY]-x(3)-[STDN]-x(2)-[ST]-H.

NAME: Lipolytic enzymes "G-D-X-G" family, putative serine active site.

CONSENSUS: [LIVM]-x-[LIVMF]-[SA]-G-D-S-[CA]-G-[GA]-x-L-[CA].

NAME: Carboxylesterases type-B serine active site.

CONSENSUS: F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-S-[STAG]-G.

NAME: Carboxylesterases type-B signature 2.

CONSENSUS: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR].

NAME: Pectinesterase signature 1.

CONSENSUS: [GSTN]-x(5)-[LIVM]-x-[LIVM]-x(2)-G-x-Y-[DNK]-E-x-[LIVM]-x-[LIVM].

NAME: Pectinesterase signature 2.

CONSENSUS: G-[STAD]-[LIVMT]-D-F-I-F-G.

NAME: Peptidyl-tRNA hydrolase signature 1.

CONSENSUS: [FY]-x(2)-T-R-H-N-x-G-x(2)-[LIVMFA](2)-[DE].

NAME: Peptidyl-tRNA hydrolase signature 2.

CONSENSUS: [GS]-x(3)-H-N-G-[LIVM]-[KR]-[DNS]-[LIVMT].

NAME: Alkaline phosphatase active site.

CONSENSUS: [IV]-x-D-S-[GAS]-[GASC]-[GAST]-[GA]-T.

NAME: Histidine acid phosphatases phosphohistidine signature.

CONSENSUS: [LIVM]-x(2)-[LIVMA]-x(2)-[LIVM]-x-R-H-[GN]-x-R-x-[PAS].

NAME: Histidine acid phosphatases active site signature.

CONSENSUS: [LIVMF]-x-[LIVMFAG]-x(2)-[STAG]-H-D-[STANQ]-x-[LIVM]-x(2)-[LIVMFY]-x(2)-

CONSENSUS: [STA].

NAME: Class A bacterial acid phosphatases signature.

CONSENSUS: G-S-Y-P-S-G-H-T.

NAME: 5'-nucleotidase signature 1.

CONSENSUS: [LIVM]-x-[LIVM](2)-[HEA]-[TI]-x-D-x-H-[GSA]-x-[LIVMF].

NAME: 5'-nucleotidase signature 2.

CONSENSUS: [FYP]-x(4)-[LIVM]-G-N-H-E-F-[DN].

NAME: Fructose-1-6-bisphosphatase active site.

CONSENSUS: [AG]-[RK]-L-x(1,2)-[LIV]-[FY]-E-x(2)-P-[LIVM]-[GSA].

NAME: Serine/threonine specific protein phosphatases signature.

CONSENSUS: [LIVM]-R-G-N-H-E.

NAME: Protein phosphatase 2A regulatory subunit PR55 signature 1.

CONSENSUS: E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N.

NAME: Protein phosphatase 2A regulatory subunit PR55 signature 2.

CONSENSUS: N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D.

NAME: Protein phosphatase 2C signature.

CONSENSUS: [LIVMFY]-[LIVMFYA]-[GSAC]-[LIVM]-[FYC]-D-G-H-[GAV].

NAME: Tyrosine specific protein phosphatases active site.

CONSENSUS: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY].

NAME: Tyrosine specific protein phosphatases profile.

NAME: Dual specificity protein phosphatase profile.

NAME: PTP type protein phosphatase profile.

NAME: Inositol monophosphatase family signature 1.

CONSENSUS: [FWV]-x(0,1)-[LIVM]-D-P-[LIVM]-D-[SG]-[ST]-x(2)-[FY]-x-[HKRNSTY].

NAME: Inositol monophosphatase family signature 2.

CONSENSUS: [WV]-D-x-[AC]-[GSA]-[GSAPV]-x-[LIVACP]-[LIV]-[LIVAC]-x(3)-[GH]-[GA].

NAME: Prokaryotic zinc-dependent phospholipase C signature.

CONSENSUS: H-Y-x-[GT]-D-[LIVM]-[DNS]-x-P-x-H-[PA]-x-N.

NAME: Phosphatidylinositol-specific phospholipase X-box domain profile.

NAME: Phosphatidylinositol-specific phospholipase Y-box domain profile.

NAME: 3'5'-cyclic nucleotide phosphodiesterases signature.

CONSENSUS: H-D-[LIVMFY]-x-H-x-[AG]-x(2)-[NQ]-x-[LIVMFY].

NAME: cAMP phosphodiesterases class-II signature.

CONSENSUS: H-x-H-L-D-H-[LIVM]-x-[GS]-[LIVMA]-[LIVM](2)-x-S-[AP].

NAME: Sulfatases signature 1.

CONSENSUS: [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)-[TR]-G.

NAME: Sulfatases signature 2.

CONSENSUS: G-[YV]-x-[ST]-x(2)-[IVA]-G-K-x(0,1)-[FYWK]-[HL].

NAME: AP endonucleases family 1 signature 1.

CONSENSUS: [APF]-D-[LIVMF](2)-x-[LIVM]-Q-E-x-K.

NAME: AP endonucleases family 1 signature 2.

CONSENSUS: D-[ST]-[FY]-R-[KH]-x(7,8)-[FYW]-[ST]-[FYW](2).

NAME: AP endonucleases family 1 signature 3.

CONSENSUS: N-x-G-x-R-[LIVM]-D-[LIVMFYH]-x-[LV]-x-S.

NAME: AP endonucleases family 2 signature 1.

CONSENSUS: H-x(2)-Y-[LIVMF]-[IM]-N-[LIVMCA]-[AG].

NAME: AP endonucleases family 2 signature 2.

CONSENSUS: [GR]-[LIVMF]-C-[LIVM]-D-T-C-H.

NAME: AP endonucleases family 2 signature 3.

CONSENSUS: [LIVMW]-H-x-N-[DE]-[SA]-K-x(3)-G-[SA]-x(2)-D.

NAME: Deoxyribonuclease I signature 1.

CONSENSUS: [LIVM](2)-[AP]-L-H-[STA](2)-P-x(5)-E-[LIVM]-[DN]-x-L-x-[DE]-V.

NAME: Deoxyribonuclease I signature 2.

CONSENSUS: G-D-F-N-A-x-C-[SA].

NAME: Endonuclease III iron-sulfur binding region signature.

CONSENSUS: C-x(3)-[KRS]-P-[KRAGL]-C-x(2)-C-x(5)-C.

NAME: Endonuclease III family signature.

CONSENSUS: [GST]-x-[LIVMF]-P-x(5)-[LIVMW]-x(2,3)-[LI]-[PAS]-G-V-[GA]-x(3)-[GAC]-

CONSENSUS: x(3)-[LIVM]-x(2)-[SALV]-[LIVMFYW]-[GANK].

NAME: Ribonuclease II family signature.

CONSENSUS: [HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-

CONSENSUS: [RQ]-[KR]-[FY]-x-D-x(3)-[HQ].

NAME: Ribonuclease III family signature.

CONSENSUS: [DEQ]-[RQ]-[LM]-E-[FYW]-[LV]-G-D-[SAR].

NAME: Bacterial Ribonuclease P protein component signature.

CONSENSUS: [LIVMFYS]-x(2)-A-x(2)-R-[NH]-[KRQL]-[LIVM]-[KRA]-R-x-[LIVMTA]-[KR].

NAME: Ribonuclease T2 family histidine active site 1.

CONSENSUS: [FYWL]-x-[LIVM]-H-G-L-W-P.

NAME: Ribonuclease T2 family histidine active site 2.

CONSENSUS: [LIVMF]-x(2)-[HDGTY]-[EQ]-[FYW]-x-[KR]-H-G-x-C.

NAME: Pancreatic ribonuclease family signature.

CONSENSUS: C-K-x(2)-N-T-F.

NAME: DNA/RNA non-specific endonucleases active site.

CONSENSUS: D-R-G-H-[QIL]-x(3)-A.

NAME: Thermonuclease family signature 1.

CONSENSUS: D-G-D-T-[LIVM]-x-[LIVMC]-x(9,10)-R-[LIVM]-x(2)-[LIVM]-D-x-P-E.

NAME: Thermonuclease family signature 2.

CONSENSUS: D-[KR]-Y-[GQ]-R-x-[LV]-[GA]-x-[IV]-[FYW].

NAME: Beta-amylase active site 1.

CONSENSUS: H-x-C-G-G-N-V-G-D.

NAME: Beta-amylase active site 2.

CONSENSUS: G-x-[SA]-G-E-[LIVM]-R-Y-P-S-Y.

NAME: Glucoamylase active site region signature.

CONSENSUS: [STN]-[GP]-x(1,2)-[DE]-x-W-E-E-x(2)-[GS].

NAME: Polygalacturonase active site.

CONSENSUS: [GSDENKRH]-x(2)-[VMFC]-x(2)-[GS]-H-G-[LIVMAG]-x(1,2)-[LIVM]-G-S.

NAME: Clostridium cellulosome enzymes repeated domain signature.

CONSENSUS: D-[LIVMFY]-[DNV]-x-[DNS]-x(2)-[LIVM]-[DN]-[SALM]-x-D-x(3)-[LIVMF]-x-

CONSENSUS: [RKS]-x-[LIVMF].

NAME: Chitinases family 18 active site.

CONSENSUS: [LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E.

NAME: Chitinases family 19 signature 1.

CONSENSUS: C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F-[GSA].

NAME: Chitinases family 19 signature 2.

CONSENSUS: [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM].

NAME: Alpha-lactalbumin / lysozyme C signature.

CONSENSUS: C-x(3)-C-x(2)-[LMF]-x(3)-[DEN]-[LI]-x(5)-C.

NAME: Alpha-galactosidase signature.

CONSENSUS: G-[LIVMFY]-x(2)-[LIVMFY]-x-[LIVM]-D-D-x-W-x(3,4)-R-[DNSF].

NAME: Trehalase signature 1.

CONSENSUS: P-G-G-R-F-x-E-x-Y-x-W-D-x-Y.

NAME: Trehalase signature 2.

CONSENSUS: Q-W-D-x-P-x-[GA]-W-[PA]-P.

NAME: Alpha-L-fucosidase putative active site.

CONSENSUS: P-x(2)-L-x(3)-K-W-E-x-C.

NAME: Glycosyl hydrolases family 1 active site.

CONSENSUS: [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN].

NAME: Glycosyl hydrolases family 1 N-terminal signature.

CONSENSUS: F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA].

NAME: Glycosyl hydrolases family 2 signature 1.

CONSENSUS: N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYW](2)-x(3)-[DN]-x(2)-

CONSENSUS: G-[LIVMFYW](4).

NAME: Glycosyl hydrolases family 2 acid/base catalyst.

CONSENSUS: [DENQF]-[KRVW]-N-H-[AP]-[SAC]-[LIVMF](3)-W-[GS]-x(2,3)-N-E.

NAME: Glycosyl hydrolases family 3 active site.

CONSENSUS: [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-

CONSENSUS: [SGADNI].

NAME: Glycosyl hydrolases family 5 signature.

CONSENSUS: [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]-[RHDNSTLIVFY].

NAME: Glycosyl hydrolases family 6 signature 1.

CONSENSUS: V-x-Y-x(2)-P-x-R-D-C-[GSAF]-x(2)-[GSA](2)-x-G.

NAME: Glycosyl hydrolases family 6 signature 2.

CONSENSUS: [LIVMYA]-[LIVA]-[LIVT]-[LIV]-E-P-D-[SAL]-[LI]-[PSAG].

NAME: Glycosyl hydrolases family 8 signature.

CONSENSUS: A-[ST]-D-[AG]-D-x(2)-[IM]-A-x-[SA]-[LIVM]-[LIVMG]-x-A-x(3)-[FW].

NAME: Glycosyl hydrolases family 9 active sites signature 1.

CONSENSUS: [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R.

NAME: Glycosyl hydrolases family 9 active sites signature 2.

CONSENSUS: [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA].

NAME: Glycosyl hydrolases family 10 active site.

CONSENSUS: [GTA]-x(2)-[LIVN]-x-[IVMF]-[ST]-E-[LIY]-[DN]-[LIVMF].

NAME: Glycosyl hydrolases family 11 active site signature 1.

CONSENSUS: [PSA]-[LQ]-x-E-Y-Y-[LIVM](2)-[DE]-x-[FYWHN].

NAME: Glycosyl hydrolases family 11 active site signature 2.

CONSENSUS: [LIVMF]-x(2)-E-[AG]-[YWG]-[QRFGS]-[SG]-[STAN]-G-x-[SAF].

NAME: Glycosyl hydrolases family 16 active sites.

CONSENSUS: E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA].

NAME: Glycosyl hydrolases family 17 signature.

CONSENSUS: [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ].

NAME: Glycosyl hydrolases family 25 active sites signature.

CONSENSUS: D-[LIVM]-x(3)-[NQ]-[PG]-x(9,10)-G-x(4)-[LIVMFY](2)-K-x-[ST]-E-[GS]-x(2)-

CONSENSUS: Y-x-[DN].

NAME: Glycosyl hydrolases family 31 active site.

CONSENSUS: [GF]-[LIVMF]-W-x-D-M-[NSA]-E.

NAME: Glycosyl hydrolases family 31 signature 2.

CONSENSUS: G-[AV]-D-[LIVMT]-C-G-[FY]-x(3)-[ST]-x(3)-L-C-x-R-W-x(2)-[LV]-[GS]-[SA]-

CONSENSUS: F-x-P-F-x-R-[DN].

NAME: Glycosyl hydrolases family 32 active site.

CONSENSUS: H-x(2)-P-x(4)-[LIVM]-N-D-P-N-G.

NAME: Glycosyl hydrolases family 35 putative active site.

CONSENSUS: G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY].

NAME: Glycosyl hydrolases family 39 active site.

CONSENSUS: W-x-F-E-x-W-N-E-P-[DN].

NAME: Glycosyl hydrolases family 45 active site.

CONSENSUS: [STA]-T-R-Y-[FYW]-D-x(5)-[CA].

NAME: Prokaryotic transglycosylases signature.

CONSENSUS: [LIVM]-x(3)-E-S-x(3)-[AP]-x(3)-S-x(5)-G-[LIVM]-[LIVMFYW]-x-[LIVMFYW]-

CONSENSUS: x(4)-[SAG].

NAME: Inosine-uridine preferring nucleoside hydrolase family signature.

CONSENSUS: D-x-D-[PT]-[GA]-x-D-D-[TAV]-[VI]-A.

NAME: Alkylbase DNA glycosidases alkA family signature.

CONSENSUS: G-I-G-x-W-[ST]-[AV]-x-[LIVMFY](2)-x-[LIVM]-x(8)-[MF]-x(2)-[ED]-D.

NAME: Formamidopyrimidine-DNA glycosylase signature.

CONSENSUS: C-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C-x(2)-C-Q.

NAME: Uracil-DNA glycosylase signature.

CONSENSUS: [KR]-[LIV]-[LIVC]-[LIVM]-x-G-[QI]-D-P-Y.

NAME: S-adenosyl-L-homocysteine hydrolase signature 1.

CONSENSUS: [CS]-N-x-[FYL]-S-[ST]-[QA]-[DEN]-x-[AV](2)-A-A-[LIV]-[SAV].

NAME: S-adenosyl-L-homocysteine hydrolase signature 2.

CONSENSUS: G-K-x(3)-[LIV]-x-G-Y-G-x-V-G-[KR]-G-x-A.

NAME: Cytosol aminopeptidase signature.

CONSENSUS: N-T-D-A-E-G-R-L.

NAME: Aminopeptidase P and proline dipeptidase signature.

CONSENSUS: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE].

NAME: Methionine aminopeptidase subfamily 1 signature.

CONSENSUS: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]-[YWV].

NAME: Methionine aminopeptidase subfamily 2 signature.

CONSENSUS: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-[DN].

NAME: Renal dipeptidase active site.

CONSENSUS: [LIVM]-E-G-[GA]-x(2)-[LIVMF]-x(6)-L-x(3)-Y-x(2)-G-[LIVM]-R.

NAME: Serine carboxypeptidases, serine active site.

CONSENSUS: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS].

NAME: Serine carboxypeptidases, histidine active site.

CONSENSUS: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x-[IVAQ]-P-x(3)-

CONSENSUS: [PSA].

NAME: Zinc carboxypeptidases, zinc-binding region 1 signature.

CONSENSUS: [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-

CONSENSUS: [LIVMFYTA].

NAME: Zinc carboxypeptidases, zinc-binding region 2 signature.

CONSENSUS: H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW].

NAME: Serine proteases, trypsin family, histidine active site.

CONSENSUS: [LIVM]-[ST]-A-[STAG]-H-C.

NAME: Serine proteases, trypsin family, serine active site.

CONSENSUS: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-

CONSENSUS: [LIVMFYSTANQH].

NAME: Serine proteases, subtilase family, aspartic acid active site.

CONSENSUS: [STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH].

NAME: Serine proteases, subtilase family, histidine active site.

CONSENSUS: H-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM].

NAME: Serine proteases, subtilase family, serine active site.

CONSENSUS: G-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG].

NAME: Serine proteases, V8 family, histidine active site.
 CONSENSUS: [ST]-G-[LIVMFYW](3)-[GN]-x(2)-T-[LIVM]-x-T-x(2)-H.

NAME: Serine proteases, V8 family, serine active site.
 CONSENSUS: T-x(2)-[GC]-[NQ]-S-G-S-x-[LIVM]-[FY].

NAME: Serine proteases, omptin family signature 1.
 CONSENSUS: W-T-D-x-S-x-H-P-x-T.

NAME: Serine proteases, omptin family signature 2.
 CONSENSUS: A-G-Y-Q-E-[ST]-R-[FYW]-S-[FYW]-[TN]-A-x-G-G-[ST]-Y.

NAME: Prolyl endopeptidase family serine active site.
 CONSENSUS: D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2).

NAME: Endopeptidase Clp serine active site.
 CONSENSUS: T-x(2)-[LIVMF]-G-x-A-[SAC]-S-[MSA]-[PAG]-[STA].

NAME: Endopeptidase Clp histidine active site.
 CONSENSUS: R-x(3)-[EAP]-x(3)-[LIVMFYT]-M-[LIVM]-H-Q-P.

NAME: ATP-dependent serine proteases, lon family, serine active site.
 CONSENSUS: D-G-[PD]-S-A-[GS]-[LIVMCA]-[TA]-[LIVM].

NAME: Eukaryotic thiol (cysteine) proteases cysteine active site.
 CONSENSUS: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV].

NAME: Eukaryotic thiol (cysteine) proteases histidine active site.
 CONSENSUS: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH].

NAME: Eukaryotic thiol (cysteine) proteases asparagine active site.
 CONSENSUS: [FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-
 CONSENSUS: [LIVMFYG]-x-[LIVMF].

NAME: Ubiquitin carboxyl-terminal hydrolase family 1 cysteine active-site.
 CONSENSUS: Q-x(3)-N-[SA]-C-G-x(3)-[LIVM](2)-H-[SA]-[LIVM]-[SA].

NAME: Ubiquitin carboxyl-terminal hydrolases family 2 signature 1.
 CONSENSUS: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-
 CONSENSUS: Q.

NAME: Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.
 CONSENSUS: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y.

NAME: Caspase family histidine active site.
 CONSENSUS: H-x(2,4)-[SC]-x(4)-[LIVMF](2)-[ST]-H-G.

NAME: Caspase family cysteine active site.
 CONSENSUS: K-P-K-[LIVMF](4)-Q-A-C-[RQG]-G.

NAME: Eukaryotic and viral aspartyl proteases active site.
 CONSENSUS: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]-x-[LIVMFSTNC]-
 CONSENSUS: x-[LIVMFGTA].

NAME: Neutral zinc metallopeptidases, zinc-binding region signature.
 CONSENSUS: [GSTALIVN]-x(2)-H-E-[LIVMFYW]-[DEHRKP]-H-x-[LIVMFYWGSPQ].

NAME: Matrixins cysteine switch.
 CONSENSUS: P-R-C-[GN]-x-P-[DR]-[LIVSAPKQ].

NAME: Insulinase family, zinc-binding region signature.
 CONSENSUS: G-x(8,9)-G-x-[STA]-H-[LIVMFY]-[LIVMC]-[DERN]-[HRKL]-[LMFAT]-x-[LFSTH]-x-
 CONSENSUS: [GSTAN]-[GST].

//

AC PS01016;
 DE Glycoprotease family signature.
 CONSENSUS: [KR]-[GSAT]-x(4)-[FYWHL]-[DQNGK]-x-P-x-[LIVMFY]-x(3)-H-x(2)-[AG]-H-
 CONSENSUS: [LIVM].

NAME: Proteasome A-type subunits signature.
 CONSENSUS: [FY]-x(4)-[STNV]-x-[FYW]-S-P-x-G-[RKH]-x(2)-Q-[LIVM]-[DE]-Y-[SAD]-x(2)-
 CONSENSUS: [SAG].

NAME: Proteasome B-type subunits signature.

CONSENSUS: [LIVMA]-[GSA]-[LIVMF]-x-[FYLVGAC]-x(2)-[GSACFY]-[LIVMSTAC](3)-[GAC]-
CONSENSUS: [GSTACV]-[DES]-x(15)-[RK]-x(12,13)-G-x(2)-[GSTA]-D.

NAME: Signal peptidases I serine active site.

CONSENSUS: [GS]-x-S-M-x-[PS]-[AT]-[LF].

NAME: Signal peptidases I lysine active site.

CONSENSUS: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY].

NAME: Signal peptidases I signature 3.

CONSENSUS: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG].

NAME: Signal peptidases II signature.

CONSENSUS: [GAF]-[GA]-[GAS]-[LIVM]-[GAS]-N-[LVMFG]-[LIVMFY]-D-R-[LIMFA].

NAME: Peptidase family U32 signature.

CONSENSUS: E-x-F-x(2)-G-[SA]-[LIVM]-C-x(4)-G-x-C-x-[LIVM]-S.

NAME: Amidases signature.

CONSENSUS: G-[GA]-S-S-[GS]-G-x-[GSA]-[GSAVY]-x-[LIVM]-[GSA]-x(6)-[GSA]-x-[GA]-x-D-
CONSENSUS: x-[GA]-x-S-[LIVM]-R-x-P-[GSAC].

NAME: Asparaginase / glutaminase active site signature 1.

CONSENSUS: [LIVM]-x(2)-T-G-G-T-[IV]-[AGS].

NAME: Asparaginase / glutaminase active site signature 2.

CONSENSUS: G-x-[LIVM]-x(2)-H-G-T-D-T-[LIVM].

NAME: Urease nickel ligands signature.

CONSENSUS: T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P.

NAME: Urease active site.

CONSENSUS: [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A.

NAME: ArgE / dapE / ACY1 / CPG2 / yscS family signature 1.

CONSENSUS: [LIV]-[GALMY]-[LIVMF]-x-[GSA]-H-x-D-[TV]-[STAV].

NAME: ArgE / dapE / ACY1 / CPG2 / yscS family signature 2.

CONSENSUS: [GSTAI]-[SANQ]-D-x-K-[GSACN]-x(2)-[LIVMA]-x(2)-[LIVMFY]-x(14,17)-[LIVM]-
CONSENSUS: x-[LIVMF]-[LIVMSTAG]-[LIVMFA]-x(2)-[DNG]-E-E-x-[GSTN].

NAME: Dihydroorotase signature 1.

CONSENSUS: D-[LIVMFYWSAP]-H-[LIVA]-H-[LIVF]-[RN]-x-[PGN].

NAME: Dihydroorotase signature 2.

CONSENSUS: [GA]-[ST]-D-x-A-P-H-x(4)-K.

NAME: Beta-lactamase class-A active site.

CONSENSUS: [FY]-x-[LIVMFY]-x-S-[TV]-x-K-x(4)-[AGLM]-x(2)-[LC].

NAME: Beta-lactamase class-C active site.

CONSENSUS: F-E-[LIVM]-G-S-[LIVMG]-[SA]-K.

NAME: Beta-lactamase class-D active site.

CONSENSUS: [PA]-x-S-[ST]-F-K-[LIV]-[PAL]-x-[STA]-[LI].

NAME: Beta-lactamases class B signature 1.

CONSENSUS: [LI]-x-[STN]-[HN]-x-H-[GSTA]-D-x(2)-G-[GP]-x(7,8)-[GS].

NAME: Beta-lactamases class B signature 2.

CONSENSUS: P-x(3)-[LIVM](2)-x-G-x-C-[LIVMF](2)-K.

NAME: Arginase family signature 1.

CONSENSUS: [LIVMF]-G-G-x-H-x-[LIVMT]-[STAV]-x-[PAG]-x(3)-[GSTA].

NAME: Arginase family signature 2.

CONSENSUS: [LIVM](2)-x-[LIVMFY]-D-[AS]-H-x-D.

NAME: Arginase family signature 3.

CONSENSUS: [ST]-[LIVMFY]-D-[LIVM]-D-x(3)-[PAQ]-x(3)-P-[GSA]-x(7)-G.

NAME: Adenosine and AMP deaminase signature.

CONSENSUS: [SA]-[LIVM]-[NGS]-[STA]-D-D-P.

NAME: Cytidine and deoxycytidylate deaminases zinc-binding region signature.

CONSENSUS: [CH]-[AGV]-E-x(2)-[LIVMFGAT]-[LIVM]-x(17,33)-P-C-x(2,8)-C-x(3)-[LIVM].

NAME: GTP cyclohydrolase 1 signature 1.

CONSENSUS: [EN]-[LIVM](2)-x(2)-[KRQN]-[DN]-[LIVM]-x(3)-[ST]-x-C-E-H-H.

NAME: GTP cyclohydrolase 1 signature 2.

CONSENSUS: [SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN].

NAME: Nitrilases / cyanide hydratase signature 1.

CONSENSUS: G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVM]-x-G-Y-P.

NAME: Nitrilases / cyanide hydratase active site signature.

CONSENSUS: G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST]-[LIVMFYS]-x-[KR].

NAME: Inorganic pyrophosphatase signature.

CONSENSUS: D-[SGDN]-D-[PE]-[LIVMF]-D-[LIVMGAC].

NAME: Acylphosphatase signature 1.

CONSENSUS: [LIV]-x-G-x-V-Q-G-V-x-[FM]-R.

NAME: Acylphosphatase signature 2.

CONSENSUS: G-[FYW]-[AVC]-[KRQAM]-N-x(3)-G-x-V-x(5)-G.

NAME: ATP synthase alpha and beta subunits signature.

CONSENSUS: P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S.

NAME: ATP synthase gamma subunit signature.

CONSENSUS: [IV]-T-x-E-x(2)-[DE]-x(3)-G-A-x-[SAKR].

NAME: ATP synthase delta (OSCP) subunit signature.

CONSENSUS: [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x-[GSA]-G-[LIVMFYGA]-

CONSENSUS: x-[LIVM]-[KRHENQ]-x-[GSEN].

NAME: ATP synthase a subunit signature.

CONSENSUS: [STAGN]-x-[STAG]-[LIVMF]-R-L-x-[SAGV]-N-[LIVMT].

NAME: ATP synthase c subunit signature.

CONSENSUS: [GSTA]-R-[NQ]-P-x(10)-[LIVMFYW](2)-x(3)-[LIVMFYW]-x-[DE].

NAME: E1-E2 ATPases phosphorylation site.

CONSENSUS: D-K-T-G-T-[LI]-[TI].

NAME: Sodium and potassium ATPases beta subunits signature 1.

CONSENSUS: [FYW]-x(2)-[FYW]-x-[FYW]-[DN]-x(6)-[LIVM]-G-R-T-x(3)-W.

NAME: Sodium and potassium ATPases beta subunits signature 2.

CONSENSUS: [RK]-x(2)-C-[RKQWI]-x(5)-L-x(2)-C-[SA]-G.

NAME: GDA1/CD39 family of nucleoside phosphatases signature.

CONSENSUS: [LIVM]-x-G-x(2)-E-G-x-[FY]-x-[FW]-[LIVA]-[TAG]-x-N-[HY].

NAME: Iodothyronine deiodinases active site.

CONSENSUS: R-P-L-V-x-N-F-G-S-[CA]-T-C-P-x-F.

NAME: Cutinase, serine active site.

CONSENSUS: P-x-[STA]-x-[LIV]-[IVT]-x-[GS]-G-Y-S-[QL]-G.

NAME: Cutinase, aspartate and histidine active sites.

CONSENSUS: C-x(3)-D-x-[IV]-C-x-G-[GST]-x(2)-[LIVM]-x(2,3)-H.

NAME: DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site.

CONSENSUS: S-[LIVMFYW]-x(5)-K-[LIVMFYWG](2)-x(3)-[LIVMFYW]-x-[CA]-x(2)-[LIVMFYWQ]-

CONSENSUS: x(2)-[RK].

NAME: Orn/Lys/Arg decarboxylases family 1 pyridoxal-P attachment site.

CONSENSUS: [STAV]-x-S-x-H-K-x(2)-[GSTAN](2)-x-[STA]-Q-[STA](2).

NAME: Orn/DAP/Arg decarboxylases family 2 pyridoxal-P attachment site.

CONSENSUS: [FY]-[PA]-x-K-[SACV]-[NHCLFW]-x(4)-[LIVMF]-[LIVMTA]-x(2)-[LIVMA]-x(3)-

CONSENSUS: [GTE].

NAME: Orn/DAP/Arg decarboxylases family 2 signature 2.
 CONSENSUS: [GS]-x(2,6)-[LIVMSCP]-x(2)-[LIVMF]-[DNS]-[LIVMCA]-G-G-G-[LIVMFY]-
 CONSENSUS: [GSTPCEQ].

NAME: Orotidine 5'-phosphate decarboxylase active site.
 CONSENSUS: [LIVMFTA]-[LIVMF]-x-D-x-K-x(2)-D-I-[GP]-x-T-[LIVMTA].

NAME: Phosphoenolpyruvate carboxylase active site 1.
 CONSENSUS: [VT]-x-T-A-H-P-T-[EQ]-x(2)-R-[KRH].

NAME: Phosphoenolpyruvate carboxylase active site 2.
 CONSENSUS: [IV]-M-[LIVM]-G-Y-S-D-S-x-K-D-[STAG]-G.

NAME: Phosphoenolpyruvate carboxykinase (GTP) signature.
 CONSENSUS: F-P-S-A-C-G-K-T-N.

NAME: Phosphoenolpyruvate carboxykinase (ATP) signature.
 CONSENSUS: L-I-G-D-D-E-H-x-W-x-[DE]-x-G-[IV]-x-N.

NAME: Uroporphyrinogen decarboxylase signature 1.
 CONSENSUS: P-x-W-x-M-R-Q-A-G-R.

NAME: Uroporphyrinogen decarboxylase signature 2.
 CONSENSUS: G-F-[STAGCV]-[STAGC]-x-P-[FYW]-T-[LV]-x(2)-Y-x(2)-[AE]-[GK].

NAME: Indole-3-glycerol phosphate synthase signature.
 CONSENSUS: [LIVMFY]-[LIVMC]-x-E-[LIVMFYC]-K-[KRSP]-[STAK]-S-P-[ST]-x(3)-[LIVMFYST].

NAME: Ribulose biphosphate carboxylase large chain active site.
 CONSENSUS: G-x-[DN]-F-x-K-x-D-E.

NAME: Fructose-bisphosphate aldolase class-I active site.
 CONSENSUS: [LIVM]-x-[LIVMFYW]-E-G-x-[LS]-L-K-P-[SN].

NAME: Fructose-bisphosphate aldolase class-II signature 1.
 CONSENSUS: [FYVM]-x(1,3)-[LIVMH]-[APN]-[LIVM]-x(1,2)-[LIVM]-H-x-D-H-[GACH].

NAME: Fructose-bisphosphate aldolase class-II signature 2.
 CONSENSUS: [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E.

NAME: Malate synthase signature.
 CONSENSUS: [KR]-[DENQ]-H-x(2)-G-L-N-x-G-x-W-D-Y-[LIVM]-F.

NAME: Hydroxymethylglutaryl-coenzyme A lyase active site.
 CONSENSUS: S-V-A-G-L-G-G-C-P-Y.

NAME: Hydroxymethylglutaryl-coenzyme A synthase active site.
 CONSENSUS: N-x-[DN]-[IV]-E-G-[IV]-D-x(2)-N-A-C-[FY]-x-G.

NAME: Citrate synthase signature.
 CONSENSUS: G-[FYA]-[GA]-H-x-[IV]-x(1,2)-[RKT]-x(2)-D-[PS]-R.

NAME: Alpha-isopropylmalate and homocitrate synthases signature 1.
 CONSENSUS: L-R-[DE]-G-x-Q-x(10)-K.

NAME: Alpha-isopropylmalate and homocitrate synthases signature 2.
 CONSENSUS: [LIVMFW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-[GASLI].

NAME: KDPG and KHG aldolases active site.
 CONSENSUS: G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R.

NAME: KDPG and KHG aldolases Schiff-base forming residue.
 CONSENSUS: G-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G.

NAME: Isocitrate lyase signature.
 CONSENSUS: K-[KR]-C-G-H-[LMQ].

NAME: Beta-eliminating lyases pyridoxal-phosphate attachment site.
 CONSENSUS: Y-x-D-x(3)-M-S-[GA]-K-K-D-x-[LIVM](2)-x-[LIVM]-G-G.

NAME: DNA photolyases class 1 signature 1.
 CONSENSUS: T-G-x-P-[LIVM](2)-D-A-x-M-[RA]-x-[LIVM].

NAME: DNA photolyases class 1 signature 2.

CONSENSUS: [DN]-R-x-R-[LIVM](2)-x-[STA](2)-F-[LIVMFA]-x-K-x-L-x(2,3)-W-[KRQ].
 NAME: DNA photolyases class 2 signature 1.
 CONSENSUS: F-x-E-E-x-[LIVM](2)-R-R-E-L-x(2)-N-F.
 NAME: DNA photolyases class 2 signature 2.
 CONSENSUS: G-x-H-D-x(2)-W-x-E-R-x-[LIVM]-F-G-K-[LIVM]-R-[FY]-M-N.
 NAME: Eukaryotic-type carbonic anhydrases signature.
 CONSENSUS: S-E-H-x-[LIVM]-x(4)-[FYH]-x(2)-E-[LIVM]-H-[LIVMFA](2).
 NAME: Prokaryotic-type carbonic anhydrases signature 1.
 CONSENSUS: C-[SA]-D-S-R-[LIVM]-x-[AP].
 NAME: Prokaryotic-type carbonic anhydrases signature 2.
 CONSENSUS: [EQ]-Y-A-[LIVM]-x(2)-[LIVM]-x(4)-[LIVMF](3)-x-G-H-x(2)-C-G.
 NAME: Fumarate lyases signature.
 CONSENSUS: G-S-x(2)-M-x(2)-K-x-N.
 NAME: Aconitase family signature 1.
 CONSENSUS: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-[LIVMA].
 NAME: Aconitase family signature 2.
 CONSENSUS: G-x(2)-[LIVWPQ]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA].
 NAME: Dihydroxy-acid and 6-phosphogluconate dehydratases signature 1.
 CONSENSUS: C-D-K-x(2)-P-[GA]-x(3)-[GA].
 NAME: Dihydroxy-acid and 6-phosphogluconate dehydratases signature 2.
 CONSENSUS: [SA]-L-[LIVM]-T-D-[GA]-R-[LIVMF]-S-[GA]-[GAV]-[ST].
 NAME: Dehydroquinase class I active site.
 CONSENSUS: D-[LIVM]-[DE]-[LIVN]-x(18,20)-[LIVM](2)-x-[SC]-[NHY]-H-[DN].
 NAME: Dehydroquinase class II signature.
 CONSENSUS: [LIVM]-[NQ]-G-P-N-[LV]-x(2)-L-G-x-R-[QED]-P-x(2)-[FY]-G.
 NAME: Enolase signature.
 CONSENSUS: [LIV](3)-K-x-N-Q-I-G-[ST]-[LIV]-[ST]-[DE]-[STA].
 NAME: Serine/threonine dehydratases pyridoxal-phosphate attachment site.
 CONSENSUS: [DESH]-x(4,5)-[STVG]-x-[AS]-[FYI]-K-[DLIFSA]-[RVMF]-[GA]-[LIVMGA].
 NAME: Enoyl-CoA hydratase/isomerase signature.
 CONSENSUS: [LIVM]-[STA]-x-[LIVM]-[DENQRHSTA]-G-x(3)-[AG](3)-x(4)-[LIVMST]-x-[CSTA]-[DQHP]-[LIVMFY].
 NAME: Imidazoleglycerol-phosphate dehydratase signature 1.
 CONSENSUS: [LIVMY]-[DE]-x-H-H-x(2)-E-x(2)-[GCA]-[LIVM]-[STAC]-[LIVM].
 NAME: Imidazoleglycerol-phosphate dehydratase signature 2.
 CONSENSUS: G-x-[DN]-x-H-H-x(2)-E-[STAGC]-x-[FY]-K.
 NAME: Tryptophan synthase alpha chain signature.
 CONSENSUS: [LIVM]-E-[LIVM]-G-x(2)-[FYC]-[ST]-[DE]-[PA]-[LIVMY]-[AGLI]-[DE]-G.
 NAME: Tryptophan synthase beta chain pyridoxal-phosphate attachment site.
 CONSENSUS: [LIVM]-x-H-x-G-[STA]-H-K-x-N.
 NAME: Delta-aminolevulinic acid dehydratase active site.
 CONSENSUS: G-x-D-x-[LIVM](2)-[IV]-K-P-[GSA]-x(2)-Y.
 NAME: Urocanase active site.
 CONSENSUS: F-Q-G-L-P-x-R-I-C-W.
 NAME: Prephenate dehydratase signature 1.
 CONSENSUS: [FY]-x-[LIVM]-x(2)-[LIVM]-x(5)-[DN]-x(5)-T-R-F-[LIVMW]-x-[LIVM].
 NAME: Prephenate dehydratase signature 2.
 CONSENSUS: [LIVM]-[ST]-[KR]-[LIVM]-E-[ST]-R-P.
 NAME: Dihydrodipicolinate synthetase signature 1.

CONSENSUS: [GSA]-[LIVM]-[LIVMFY]-x(2)-G-[ST]-[TG]-G-E-[GASNF]-x(6)-[EQ].

NAME: Dihydrodipicolinate synthetase signature 2.

CONSENSUS: Y-[DNS]-[LIVMF]-P-x(2)-[ST]-x(3)-[LIVM]-x(13,14)-[LIVM]-x-[SGA]-[LIVMF]-

CONSENSUS: K-[DEQAF]-[STAC].

NAME: RsaA family of pseudouridine synthase signature.

CONSENSUS: G-R-L-D-x(2)-[ST]-x-G-[LIVMF](4)-[ST]-[DNT].

NAME: Cysteine synthase/cystathionine beta-synthase P-phosphate attachment site.

CONSENSUS: K-x-E-x(3)-[PA]-[STAGC]-x-S-[IVAP]-K-x-R-x-[STAG]-x(2)-[LIVM].

NAME: Phenylalanine and histidine ammonia-lyases signature.

CONSENSUS: G-[STG]-[LIVM]-[STG]-[AC]-S-G-[DH]-L-x-P-L-[SA]-x(2)-[SA].

NAME: Porphobilinogen deaminase cofactor-binding site.

CONSENSUS: E-R-x-[LIVMFA]-x(3)-[LIVMF]-x-G-[GSA]-C-x-[IVT]-P-[LIVMF]-[GSA].

NAME: Cys/Met metabolism enzymes pyridoxal-phosphate attachment site.

CONSENSUS: [DQ]-[LIVMF]-x(3)-[STAGC]-[STAGCI]-T-K-[FYWQ]-[LIVMF]-x-G-[HQ]-[SGNH].

NAME: Glyoxalase I signature 1.

CONSENSUS: [HQ]-[IVT]-x-[LIVFY]-x-[IV]-x(5)-[STA]-x(2)-F-[YM]-x(2,3)-[LMF]-G-[LMF].

NAME: Glyoxalase I signature 2.

CONSENSUS: G-[NTKQ]-x(0,5)-[GA]-[LVFY]-[GH]-H-[IVF]-[CGA]-x-[STAGL]-x(2)-[DNC].

NAME: Cytochrome c and c1 heme lyases signature 1.

CONSENSUS: H-N-x(2)-N-E-x(2)-W-[NQKR]-x(4)-W-E.

NAME: Cytochrome c and c1 heme lyases signature 2.

CONSENSUS: P-F-D-R-H-D-W.

NAME: Adenylate cyclases class-I signature 1.

CONSENSUS: E-Y-F-G-[SA](2)-L-W-x-L-Y-K.

NAME: Adenylate cyclases class-I signature 2.

CONSENSUS: Y-R-N-x-W-[NS]-E-[LIVM]-R-T-L-H-F-x-G.

NAME: Guanylate cyclases signature.

CONSENSUS: G-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-[DNT]-[IV]-

CONSENSUS: [DNTA]-x(5)-[DE].

NAME: Chorismate synthase signature 1.

CONSENSUS: G-E-S-H-[GC]-x(2)-[LIVM]-[GTV]-x-[LIVM](2)-[DE]-G-x-[PV].

NAME: Chorismate synthase signature 2.

CONSENSUS: [GE]-R-[SA](2)-[SAG]-R-[EV]-[ST]-x(2)-[RH]-V-x(2)-G.

NAME: Chorismate synthase signature 3.

CONSENSUS: R-[SH]-D-[PSV]-[CSAV]-x(4)-[GAI]-x-[IVGSP]-[LIVM]-x-E-[STAH]-[LIVM].

NAME: 6-pyruvoyl tetrahydropterin synthase signature 1.

CONSENSUS: C-N-N-x(2)-G-H-G-H-N-Y.

NAME: 6-pyruvoyl tetrahydropterin synthase signature 2.

CONSENSUS: D-H-K-N-L-D-x-D.

NAME: Ferrochelatase signature.

CONSENSUS: [LIVMF](2)-x-S-x-H-[GS]-[LIVM]-P-x(4,5)-[DENQKR]-x-G-D-x-Y.

NAME: Alanine racemase pyridoxal-phosphate attachment site.

CONSENSUS: V-x-K-A-[DN]-[GA]-Y-G-H-G.

NAME: Aspartate and glutamate racemases signature 1.

CONSENSUS: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK].

NAME: Aspartate and glutamate racemases signature 2.

CONSENSUS: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM].

NAME: Mandelate racemase / muconate lactonizing enzyme family signature 1.

CONSENSUS: A-x-[SAG](2)-[LIVM]-[DE]-x-A-x(2)-D-x(2)-[GA]-[KR].

NAME: Mandelate racemase / muconate lactonizing enzyme family signature 2.

CONSENSUS: G-x(7)-D-x(9)-A-x(14)-[LIVM]-E-[DENQ]-P-x(4)-[DENQ].

NAME: Ribulose-phosphate 3-epimerase family signature 1.

CONSENSUS: [LIVMF]-H-[LIVMFY]-D-[LIVM]-x-D-x(1,2)-[FY]-[LIVM]-x-N-x-[STAV].

NAME: Ribulose-phosphate 3-epimerase family signature 2.

CONSENSUS: [LIVMA]-x-[LIVM]-M-[ST]-[VS]-x-P-x(3)-G-Q-x-F-x(6)-[NK]-[LIVMC].

NAME: Aldose 1-epimerase putative active site.

CONSENSUS: [NS]-x-T-N-H-x-Y-[FW]-N-[LI].

NAME: Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature.

CONSENSUS: [FY]-x(2)-[STCNLV]-x-F-H-[RH]-[LIVMN]-[LIVM]-x(2)-F-[LIVM]-x-Q-[AG]-G.

NAME: Cyclophilin-type peptidyl-prolyl cis-trans isomerase profile.

NAME: FKBP-type peptidyl-prolyl cis-trans isomerase signature 1.

CONSENSUS: [LIVMC]-x-[YF]-x-[GVL]-x(1,2)-[LFT]-x(2)-G-x(3)-[DE]-[STAEQK]-[STAN].

NAME: FKBP-type peptidyl-prolyl cis-trans isomerase signature 2.

CONSENSUS: [LIVMFY]-x(2)-[GA]-x(3,4)-[LIVMF]-x(2)-[LIVMFHK]-x(2)-G-x(4)-[LIVMF]-

CONSENSUS: x(3)-[PSGAQ]-x(2)-[AG]-[FY]-G.

NAME: FKBP-type peptidyl-prolyl cis-trans isomerase domain profile.

NAME: PpiC-type peptidyl-prolyl cis-trans isomerase signature.

CONSENSUS: F-[GSADEI]-x-[LVAQ]-A-x(3)-[ST]-x(3,4)-[STQ]-x(3,5)-[GER]-G-x-[LIVM]-

CONSENSUS: [GS].

NAME: Triosephosphate isomerase active site.

CONSENSUS: [AV]-Y-E-P-[LIVM]-W-[SA]-I-G-T-[GK].

NAME: Xylose isomerase signature 1.

CONSENSUS: [LI]-E-P-K-P-x(2)-P.

NAME: Xylose isomerase signature 2.

CONSENSUS: [FL]-H-D-x-D-[LIV]-x-[PD]-x-[GDE].

NAME: Phosphomannose isomerase type I signature 1.

CONSENSUS: Y-x-D-x-N-H-K-P-E.

NAME: Phosphomannose isomerase type I signature 2.

CONSENSUS: H-A-Y-[LIVM]-x-G-x(2)-[LIVM]-E-x-M-A-x-S-D-N-x-[LIVM]-R-A-G-x-T-P-K.

NAME: Phosphoglucose isomerase signature 1.

CONSENSUS: [DENS]-x-[LIVM]-G-G-R-[FY]-S-[LIVMT]-x-[STA]-[PSAC]-[LIVMA]-G.

NAME: Phosphoglucose isomerase signature 2.

CONSENSUS: [GS]-x-[LIVM]-[LIVMFYW]-x(4)-[FY]-[DN]-Q-x-G-V-E-x(2)-K.

NAME: Glucosamine/galactosamine-6-phosphate isomerases signature.

CONSENSUS: [LIVM]-x(3)-G-x-[LIT]-x-[LIV]-x-[LIVM]-x-G-[LIVM]-G-x-[DEN]-G-H.

NAME: Phosphoglycerate mutase family phosphohistidine signature.

CONSENSUS: [LIVM]-x-R-H-G-[EQ]-x(3)-N.

NAME: Phosphoglucomutase and phosphomannomutase phosphoserine signature.

CONSENSUS: [GSA]-[LIVM]-x-[LIVM]-[ST]-[PGA]-S-H-x-P-x(4)-[GNHE].

NAME: Methylmalonyl-CoA mutase signature.

CONSENSUS: R-I-A-R-N-[TQ]-x(2)-[LIVMFY](2)-x-[EQ]-E-x(4)-[KRN]-x(2)-D-P-x-[GSA]-

CONSENSUS: G-S.

NAME: Terpene synthases signature.

CONSENSUS: [DE]-G-S-W-x-G-x-W-[GA]-[LIVM]-x-[FY]-x-Y-[GA].

NAME: Eukaryotic DNA topoisomerase I active site.

CONSENSUS: [DEN]-x(6)-[GS]-[IT]-S-K-x(2)-Y-[LIVM]-x(3)-[LIVM].

NAME: Prokaryotic DNA topoisomerase I active site.

CONSENSUS: [EQ]-x-L-Y-[DEQT]-x(3,12)-[LI]-[ST]-Y-x-R-[ST]-[DEQS].

NAME: DNA topoisomerase II signature.

CONSENSUS: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG].

NAME: Aminoacyl-transfer RNA synthetases class-I signature.

CONSENSUS: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-
CONSENSUS: [LIVMFYSTAGPC].

NAME: Aminoacyl-transfer RNA synthetases class-II signature 1.

CONSENSUS: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE].

NAME: Aminoacyl-transfer RNA synthetases class-II signature 2.

CONSENSUS: [GSTALVF]-[DENQHRKP]-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY].

NAME: WHEP-TRS domain signature.

CONSENSUS: [QY]-G-[DNEA]-x-[LIV]-[KR]-x(2)-K-x(2)-[KRNG]-[AS]-x(4)-[LIV]-[DENK]-
CONSENSUS: x(2)-[IV]-x(2)-L-x(3)-K.

NAME: ATP-citrate lyase / succinyl-CoA ligases family signature 1.

CONSENSUS: S-[KR]-S-G-[GT]-[LIVM]-[GST]-x-[EQ]-x(8,10)-G-x(4)-[LIVM]-[GA]-[LIVM]-G-
CONSENSUS: G-D.

NAME: ATP-citrate lyase / succinyl-CoA ligases family active site.

CONSENSUS: G-x(2)-A-x(4,7)-[RQT]-[LIVMF]-G-H-[AS]-[GH].

NAME: ATP-citrate lyase / succinyl-CoA ligases family signature 3.

CONSENSUS: G-x-[IV]-x(2)-[LIVMF]-x-[NA]-G-[GA]-G-[LA]-[STAV]-x(4)-D-x-[LIVM]-x(3)-
CONSENSUS: G-[GRE].

NAME: Glutamine synthetase signature 1.

CONSENSUS: [FYWL]-D-G-S-S-x(6,8)-[DENQSTAK]-[SA]-[DE]-x(2)-[LIVMFY].

NAME: Glutamine synthetase putative ATP-binding region signature.

CONSENSUS: K-P-[LIVMFYA]-x(3,5)-[NPAT]-G-[GSTAN]-G-x-H-x(3)-S.

NAME: Glutamine synthetase class-I adenylation site.

CONSENSUS: K-[LIVM]-x(5)-[LIVMA]-D-[RK]-[DN]-[LI]-Y.

NAME: D-alanine--D-alanine ligase signature 1.

CONSENSUS: H-G-x(2)-G-E-D-G-x-[LIVMA]-[QSA]-[GSA].

NAME: D-alanine--D-alanine ligase signature 2.

CONSENSUS: [LIV]-x(3)-[GA]-x-[GSAIV]-R-[LIVCA]-D-[LIVMF](2)-x(7,9)-[LI]-x-E-
CONSENSUS: [LIVA]-N-[STP]-x-P-[GA].

NAME: SAICAR synthetase signature 1.

CONSENSUS: [LIVMF](2)-P-[LIVM]-E-x-[LIVM]-[LIVMCA]-R-x(3)-[TA]-G-S.

NAME: SAICAR synthetase signature 2.

CONSENSUS: [LIVM]-[LIVMA]-D-x-K-[LIVMFY]-E-F-G.

NAME: Folylpolylglutamate synthase signature 1.

CONSENSUS: [LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)-[LIVM](2)-x(3)-[GSK].

NAME: Folylpolylglutamate synthase signature 2.

CONSENSUS: [LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2).

NAME: Ubiquitin-activating enzyme signature 1.

CONSENSUS: K-A-C-S-G-K-F-x-P.

NAME: Ubiquitin-activating enzyme active site.

CONSENSUS: P-[LIVM]-C-T-[LIVM]-[KRH]-x-[FT]-P.

NAME: Ubiquitin-conjugating enzymes active site.

CONSENSUS: [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x-[LIV].

NAME: Formate--tetrahydrofolate ligase signature 1.

CONSENSUS: G-[LIVM]-K-G-G-A-A-G-G-G-Y.

NAME: Formate--tetrahydrofolate ligase signature 2.

CONSENSUS: V-A-T-[IV]-R-A-L-K-x-[HN]-G-G.

NAME: Adenylosuccinate synthetase GTP-binding site.

CONSENSUS: Q-W-G-D-E-G-K-G.

NAME: Adenylosuccinate synthetase active site.

CONSENSUS: G-I-[GR]-P-x-Y-x(2)-K-x(2)-R.

NAME: Argininosuccinate synthase signature 1.
 CONSENSUS: A-[FY]-S-G-G-L-D-T-S.

NAME: Argininosuccinate synthase signature 2.
 CONSENSUS: G-x-T-x-K-G-N-D-x(2)-R-F.

NAME: Phosphoribosylglycinamide synthetase signature.
 CONSENSUS: R-F-G-D-P-E-x-[QM].

NAME: Carbamoyl-phosphate synthase subdomain signature 1.
 CONSENSUS: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-[SG]-G-x-[AG].

NAME: Carbamoyl-phosphate synthase subdomain signature 2.
 CONSENSUS: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC].

NAME: ATP-dependent DNA ligase AMP-binding site.
 CONSENSUS: [EDQH]-x-K-x-[DN]-G-x-R-[GACIVM].

NAME: ATP-dependent DNA ligase signature 2.
 CONSENSUS: E-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-[KRH]-x(3,5)-K-
 CONSENSUS: [LIVMFY]-K.

NAME: NAD-dependent DNA ligase signature 1.
 CONSENSUS: K-[LIVM]-D-G-[LIVM]-[SA]-x(4)-Y-x(2)-G-x-L-x(4)-[ST]-R-G-[DN]-G-x(2)-G-
 CONSENSUS: [DE]-[DENL].

NAME: NAD-dependent DNA ligase signature 2.
 CONSENSUS: [IV]-G-[KR]-[ST]-G-x-[LIVM]-[STNK]-x-[VT]-x(2)-L-x-[PS]-V.

NAME: RNA 3'-terminal phosphate cyclase signature.
 CONSENSUS: [RH]-G-x(2)-P-x-G(3)-x-[LIV].

NAME: Lipoate-protein ligase B signature.
 CONSENSUS: R-G-G-x(2)-T-[FYW]-H-x(2)-[GH]-Q-x-[LIV]-x-Y.

NAME: Isopenicillin N synthetase signature 1.
 CONSENSUS: [RK]-x-[STA]-x(2)-S-x-C-Y-[SL].

NAME: Isopenicillin N synthetase signature 2.
 CONSENSUS: [LIVM](2)-x-C-G-[STA]-x(2)-[STAG]-x(2)-T-x-[DNG].

NAME: Site-specific recombinases active site.
 CONSENSUS: Y-[LIVAC]-R-[VA]-S-[ST]-x(2)-Q.

NAME: Site-specific recombinases signature 2.
 CONSENSUS: G-[DE]-x(2)-[LIVM]-x(3)-[LIVM]-[DT]-R-[LIVM]-[GSA].

NAME: Transposases, Mutator family, signature.
 CONSENSUS: D-x(3)-G-[LIVMF]-x(6)-[STAV]-[LIVMFYW]-[PT]-x-[STAV]-x(2)-[QR]-x-C-x(2)-
 CONSENSUS: H.

NAME: Transposases, IS30 family, signature.
 CONSENSUS: R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K.

NAME: Autoinducers synthetases family signature.
 CONSENSUS: [LMFY]-R-x(3)-F-x(2)-[KR]-x(2)-W-x-[LIVM]-x(6,9)-E-x-D-x-[FY]-D.

NAME: Thiamine pyrophosphate enzymes signature.
 CONSENSUS: [LIVMF]-[GSA]-x(5)-P-x(4)-[LIVMFYW]-x-[LIVMF]-x-G-D-[GSA]-[GSAC].

NAME: Biotin-requiring enzymes attachment site.
 CONSENSUS: [GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]-x(3)-[LIVM]-x-
 CONSENSUS: [SAV].

NAME: 2-oxo acid dehydrogenases acyltransferase component lipoyl binding site.
 CONSENSUS: [GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]-[STAVQDN]-
 CONSENSUS: x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY].

NAME: Putative AMP-binding domain signature.
 CONSENSUS: [LIVMFY]-x(2)-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-x-[PASLIVM]-[KR].

NAME: Molybdenum cofactor biosynthesis proteins signature 1.
 CONSENSUS: [LIVM](3)-[LIT](2)-G-G-T-G-x(4)-D.

NAME: Molybdenum cofactor biosynthesis proteins signature 2.

CONSENSUS: S-x-[GS]-x(2)-D-x(5)-[LIVW]-x(10,12)-[LIV]-x(2)-[KR]-P-G-[KRL]-P-x(2)-[LIVMF]-[GA].

NAME: moaA / nifB / pqqE family signature.

CONSENSUS: [LIV]-x(3)-C-[NP]-[LIVMF]-[QRS]-C-x-[FYM]-C.

NAME: Radical activating enzymes signature.

CONSENSUS: [GV]-x-G-x-[KR]-x(3)-F-x(2)-G-x(0,1)-C-x(3)-C-x(2)-C-x-[NL].

NAME: Tpx family signature.

CONSENSUS: S-x-D-L-P-F-A-x(2)-[KR]-[FW]-C.

NAME: Cytochrome c family heme-binding site signature.

CONSENSUS: C-[CPWHF]-[CPWR]-C-H-[CFYW].

NAME: Cytochrome b5 family, heme-binding domain signature.

CONSENSUS: [FY]-[LIVMK]-x(2)-H-P-[GA]-G.

NAME: Cytochrome b/b6 heme-ligand signature.

CONSENSUS: [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H.

NAME: Cytochrome b/b6 Qo site signature.

CONSENSUS: P-[DE]-W-[FY]-[LFY](2).

NAME: Cytochrome b559 subunits heme-binding site signature.

CONSENSUS: [LIV]-x-[ST]-[LIVF]-R-[FYW]-x(2)-[IV]-H-[STGA]-[LIV]-[STGA]-[IV]-P.

NAME: Nickel-dependent hydrogenases b-type cytochrome subunit signature 1.

CONSENSUS: R-[LIVMFYW]-x-H-W-[LIVM]-x(2)-[LIVMF]-[STAC]-[LIVM]-x(2)-L-x-[LIVM]-T-G.

NAME: Nickel-dependent hydrogenases b-type cytochrome subunit signature 2.

CONSENSUS: [RHT]-[STA]-[LIVMFYW]-H-[RHT]-[LIVM]-x(2)-W-x-[LIVMF]-x(2)-F-x(3)-H.

NAME: Succinate dehydrogenase cytochrome b subunit signature 1.

CONSENSUS: R-P-[LIVMT]-x(3)-[LIVM]-x(6)-[LIVMWPK]-x(4)-S-x(2)-H-R-x-[ST].

NAME: Succinate dehydrogenase cytochrome b subunit signature 2.

CONSENSUS: H-x(3)-[GA]-[LIVMT]-R-[HF]-[LIVMF]-x-[FYWM]-D-x-[GVA].

NAME: Thioredoxin family active site.

CONSENSUS: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT].

NAME: Glutaredoxin active site.

CONSENSUS: [LIVD]-[FYSA]-x(4)-C-[PV]-[FYW]-C-x(2)-[TAV]-x(2,3)-[LIV].

NAME: Type-1 copper (blue) proteins signature.

CONSENSUS: [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ].

NAME: 2Fe-2S ferredoxins, iron-sulfur binding region signature.

CONSENSUS: C-[C]-[C]-[GA]-[C]-C-[GAST]-[CPDEKRHFYW]-C.

NAME: Adrenodoxin family, iron-sulfur binding region signature.

CONSENSUS: C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR].

NAME: 4Fe-4S ferredoxins, iron-sulfur binding region signature.

CONSENSUS: C-x(2)-C-x(2)-C-x(3)-C-[PEG].

NAME: High potential iron-sulfur proteins signature.

CONSENSUS: C-x(6,9)-[LIVM]-x(3)-G-[YW]-C-x(2)-[FYW].

NAME: Rieske iron-sulfur protein signature 1.

CONSENSUS: C-[TK]-H-L-G-C-[LIVT].

NAME: Rieske iron-sulfur protein signature 2.

CONSENSUS: C-P-C-H-x-[GSA].

NAME: Flavodoxin signature.

CONSENSUS: [LIV]-[LIVFY]-[FY]-x-[ST]-x(2)-[AGC]-x-T-x(3)-A-x(2)-[LIV].

NAME: Rubredoxin signature.

CONSENSUS: [LIVM]-x(3)-W-x-C-P-x-C-[AGD].

NAME: Electron transfer flavoprotein alpha-subunit signature.

CONSENSUS: [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)-[IV]-x-A-[IV]-N.

NAME: Electron transfer flavoprotein beta-subunit signature.

CONSENSUS: [IVA]-x-[KR]-x(2)-[DE]-[GD]-[GDE]-x(1,2)-[EQ]-x-[LIV]-x(4)-P-x-[LIVM](2)-[TAC].

NAME: Vertebrate metallothioneins signature.

CONSENSUS: C-x-C-[GSTAP]-x(2)-C-x-C-x(2)-C-x-C-x(2)-C-x-K.

NAME: Ferritin iron-binding regions signature 1.

CONSENSUS: E-x-[KR]-E-x(2)-E-[KR]-[LF]-[LIVMA]-x(2)-Q-N-x-R-x-G-R.

NAME: Ferritin iron-binding regions signature 2.

CONSENSUS: D-x(2)-[LIVMF]-[STAC]-[DH]-F-[LI]-[EN]-x(2)-[FY]-L-x(6)-[LIVM]-[KN].

NAME: Bacterioferritin signature.

CONSENSUS: <M-x-G-x(3)-V-[LIV]-x(2)-[LM]-x(3)-L-x(3)-L.

NAME: Transferrins signature 1.

CONSENSUS: Y-x(0,1)-[VAS]-V-[IVAC]-[IVA]-[IVA]-[RKH]-[RKS]-[GDENSA].

NAME: Transferrins signature 2.

CONSENSUS: Y-x-G-A-[FL]-[KRHNQ]-C-L-x(3,4)-G-[DENQ]-V-[GA]-[FYW].

NAME: Transferrins signature 3.

CONSENSUS: [DENQ]-[YF]-x-[LY]-L-C-x-[DN]-x(5,8)-[LIV]-x(4,5)-C-x(2)-A-x(4)-[HQR]-x-[LIVMFYW]-[LIVM].

NAME: Globins profile.

NAME: Protozoan/cyanobacterial globins signature.

CONSENSUS: F-[LF]-x(5)-G-[PA]-x(4)-G-[KRA]-x-[LIVM]-x(3)-H.

NAME: Plant hemoglobins signature.

CONSENSUS: [SN]-P-x-L-x(2)-H-A-x(3)-F.

NAME: Hemerythrins signature.

CONSENSUS: W-L-x-[NQ]-H-I-x(3)-D-F.

NAME: Arthropod hemocyanins / insect LSPs signature 1.

CONSENSUS: Y-[FYW]-x-E-D-[LIVM]-x(2)-N-x(6)-H-x(3)-P.

NAME: Arthropod hemocyanins / insect LSPs signature 2.

CONSENSUS: T-x(2)-R-D-P-x-[FY]-[FYW].

NAME: Heavy-metal-associated domain.

CONSENSUS: [LIVN]-x(2)-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-x(3)-[LIV]-x(9,11)-[IVA]-x-[LVFYS].

NAME: ABC transporters family signature.

CONSENSUS: [LIVMFYC]-[SA]-[SAPGLVFKQH]-G-[DENQMW]-[KRQASPLIMFW]-[KRNQSTAVM]-[KRACLVM]-[LIVMFYPAN]-[PHY]-[LIVMFW]-[SAGCLIVP]-[FYWHP]-[KRHP]-[LIVMFYWSTA].

NAME: Binding-protein-dependent transport systems inner membrane comp. sign.

CONSENSUS: [LIVMFY]-x(8)-[EQR]-[STAGV]-[STAG]-x(3)-G-[LIVMFYSTAC]-x(5)-[LIVMFYSTA]-x(4)-[LIVMFY]-[PKR].

NAME: ABC-2 type transport system integral membrane proteins signature.

CONSENSUS: [LIMST]-x(2)-[LIMW]-x(2)-[LIMCA]-[GSTC]-x-[GSAIV]-x(6)-[LIMGA]-[PGSNQ]-x(9,12)-P-[LIMFT]-x-[HRSY]-x(5)-[RQ].

NAME: Bacterial extracellular solute-binding proteins, family 1 signature.

CONSENSUS: [GAP]-[LIVMFA]-[STAVDN]-x(4)-[GSAV]-[LIVMFY](2)-Y-[ND]-x(3)-[LIVMF]-x-[KNDE].

NAME: Bacterial extracellular solute-binding proteins, family 3 signature.

CONSENSUS: G-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-[LIVMAGN].

NAME: Bacterial extracellular solute-binding proteins, family 5 signature.

CONSENSUS: [AG]-x(6,7)-[DNEG]-x(2)-[STAVE]-[LIVMFYWA]-x-[LIVMFY]-x-[LIVM]-[KR].

CONSENSUS: [KRHDE]-[GDN]-[LIVMA]-[KNGSP]-[FW].

NAME: Serum albumin family signature.

CONSENSUS: [FY]-x(6)-C-C-x(7)-C-[LFY]-x(6)-[LIVMFYW].

NAME: Transthyretin signature 1.

CONSENSUS: S-K-C-P-L-M-V-K-V-L-D-[AS]-V-R-G.

NAME: Transthyretin signature 2.

CONSENSUS: S-P-[FY]-S-[FY]-S-T-T-A-[LIVM]-V-[ST]-x-P.

NAME: Avidin / Streptavidin family signature.

CONSENSUS: [DEN]-x(2)-[KR]-[STA]-x(2)-V-G-x-[DN]-x-[FW]-T-[KR].

NAME: Eukaryotic cobalamin-binding proteins signature.

CONSENSUS: [SN]-V-D-T-[GA]-A-[LIVM]-A-x-L-A-[LIVMF]-T-C.

NAME: Lipocalin signature.

CONSENSUS: [DENG]-x-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-[CP]-G-{C}-W-[FYWLRH]-x-

CONSENSUS: [LIVMTA].

NAME: Cytosolic fatty-acid binding proteins signature.

CONSENSUS: [GSAIVK]-x-[FYW]-x-[LIVMF]-x(4)-[NHG]-[FY]-[DE]-x-[LIVMFY]-[LIVM]-x(2)-

CONSENSUS: [LIVMAKR].

NAME: Acyl-CoA-binding protein signature.

CONSENSUS: P-[STA]-x-[DEN]-x-[LIVMF]-x(2)-[LIVMFY]-Y-[GSTA]-x-[FY]-K-Q-[STA](2)-x-G.

NAME: LBP / BPI / CETP family signature.

CONSENSUS: [PA]-[GA]-[LIVMC]-x(2)-R-[IV]-[ST]-x(3)-L-x(5)-[EQ]-x(4)-[LIVM]-[EQK]-

CONSENSUS: x(8)-P.

NAME: Phosphatidylethanolamine-binding protein family signature.

CONSENSUS: [FY]-x-[LIVMF](3)-x-[DC]-P-D-x-P-[SN]-x(10)-H.

NAME: Plant lipid transfer proteins signature.

CONSENSUS: [LIVM]-[PA]-x(2)-C-x-[LIVM]-x-[LIVM]-x-[LIVMFY]-x-[LIVM]-[ST]-x(3)-

CONSENSUS: [DN]-C-x(2)-[LIVM].

NAME: Uteroglobin family signature 1.

CONSENSUS: [GA]-x(3)-I-C-P-x-[LIVMF]-x(3)-[LIVM]-[DE]-x-[LIVMF](2).

NAME: Uteroglobin family signature 2.

CONSENSUS: [DEQ]-x(4)-[SN]-x(5)-[DEQ]-x-I-x(2)-S-[PSE]-[LS]-C.

NAME: Mitochondrial energy transfer proteins signature.

CONSENSUS: P-x-[DE]-x-[LIVAT]-[RK]-x-[LRH]-[LIVMFY]-[QMAIGV].

NAME: Sugar transport proteins signature 1.

CONSENSUS: [LIVMSTAG]-[LIVMFSAG]-x(2)-[LIVMSA]-[DE]-x-[LIVMFYWA]-G-R-[RK]-x(4,6)-

CONSENSUS: [GSTA].

NAME: Sugar transport proteins signature 2.

CONSENSUS: [LIVMF]-x-G-[LIVMFA]-x(2)-G-x(8)-[LIFY]-x(2)-[EQ]-x(6)-[RK].

NAME: LacY family proton/sugar symporters signature 1.

CONSENSUS: G-[LIVM](2)-x-D-[RK]-L-G-L-[RK](2)-x-[LIVM](2)-W.

NAME: LacY family proton/sugar symporters signature 2.

CONSENSUS: P-x-[LIVMF](2)-N-R-[LIVM]-G-x-K-N-[STA]-[LIVM](3).

NAME: PTR2 family proton/oligopeptide symporters signature 1.

CONSENSUS: [GA]-[GAS]-[LIVMFYWA]-[LIVM]-[GAS]-D-x-[LIVMFYWT]-[LIVMFYW]-G-x(3)-[TAV]-

CONSENSUS: [IV]-x(3)-[GSTAV]-x-[LIVMF]-x(3)-[GA].

NAME: PTR2 family proton/oligopeptide symporters signature 2.

CONSENSUS: [FYT]-x(2)-[LMFY]-[FYV]-[LIVMFYWA]-x-[IVG]-N-[LIVMAG]-G-[GSA]-[LIMF].

NAME: Amiloride-sensitive sodium channels signature.

CONSENSUS: Y-x(2)-[EQTF]-x-C-x(2)-[GSTDNL]-C-x-[QT]-x(2)-[LIVMT]-[LIVMS]-x(2)-C-x-C.

NAME: Sodium:alanine symporter family signature.

CONSENSUS: G-G-x-[GA](2)-[LIVM]-F-W-M-W-[LIVM]-x-[STAV]-[LIVMFA](2)-G.

NAME: Sodium:dicarboxylate symporter family signature 1.
 CONSENSUS: P-x(0,1)-G-[DE]-x-[LIVMF](2)-x-[LIVM](2)-[KREQ]-[LIVM](3)-x-P.

NAME: Sodium:dicarboxylate symporter family signature 2.
 CONSENSUS: P-x-G-x-[STA]-x-[NT]-[LIVMC]-D-G-[STAN]-x-[LIVM]-[FY]-x(2)-[LIVM]-x(2)-[LIVM]-[FY]-[LI]-[SA]-Q.

NAME: Sodium:galactoside symporter family signature.
 CONSENSUS: D-x(3)-G-x(3)-[DN]-x(6,8)-G-[KH]-F-[KR]-P-[FYW]-[LIVM](2)-x-[GSTA](2).

NAME: Sodium:neurotransmitter symporter family signature 1.
 CONSENSUS: W-R-F-[GP]-Y-x(4)-N-G-G-G-x-[FY].

NAME: Sodium:neurotransmitter symporter family signature 2.
 CONSENSUS: Y-[LIVMFY]-x(2)-[SC]-[LIVMFY]-[STQ]-x(2)-L-P-W-x(2)-C-x(4)-N-[GST].

NAME: Sodium:solute symporter family signature 1.
 CONSENSUS: [GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-
 CONSENSUS: [SAP].

NAME: Sodium:solute symporter family signature 2.
 CONSENSUS: [GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-
 CONSENSUS: x-[LIVMG].

NAME: Sodium:sulfate symporter family signature.
 CONSENSUS: [STACP]-S-x(2)-F-x(2)-P-[LIVM]-[GSA]-x(3)-N-x-[LIVM]-V.

NAME: glpT family of transporters signature.
 CONSENSUS: R-G-x(5)-W-N-x(2)-H-N-x-G-G.

NAME: Ammonium transporters signature.
 CONSENSUS: D-[FYWS]-A-G-[GSC]-x(2)-[IV]-x(3)-[SAG](2)-x(2)-[SAG]-[LIVMF]-x(3)-
 CONSENSUS: [LIVMFYWA](2)-x-[GK]-x-R.

NAME: BCCT family of transporters signature.
 CONSENSUS: [GSDN]-W-T-[LIVM]-x-[FY]-W-x-W-W.

NAME: Flagellar motor protein motA family signature.
 CONSENSUS: A-[LMF]-x-[GAT]-T-[LIVF]-x-G-x-[LIVMF]-x(7)-P.

NAME: Formate and nitrite transporters signature 1.
 CONSENSUS: [LIVMA]-[LIVMY]-x-G-[GSTA]-[DES]-L-[FI]-[TN]-[GS].

NAME: Formate and nitrite transporters signature 2.
 CONSENSUS: [GA]-x(2)-[CA]-N-[LIVMFYW](2)-V-C-[LV]-A.

NAME: Prokaryotic sulfate-binding proteins signature 1.
 CONSENSUS: K-x-[NQEK]-[GT]-G-[DQ]-x-[LIVM]-x(3)-Q-S.

NAME: Prokaryotic sulfate-binding proteins signature 2.
 CONSENSUS: N-P-K-[ST]-S-G-x-A-R.

NAME: Sulfate transporters signature.
 CONSENSUS: P-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVMFY](3)-x(3)-[GSTA](2)-S-[KR].

NAME: Amino acid permeases signature.
 CONSENSUS: [STAGC]-G-[PAG]-x(2,3)-[LIVMFYWA](2)-x-[LIVMFYW]-x-[LIVMFWSTAGC](2)-
 CONSENSUS: [STAGC]-x(3)-[LIVMFYW]-x-[LIVMST]-x(3)-[LIMCTA]-[GA]-E-x(5)-[PSAL].

NAME: Aromatic amino acids permeases signature.
 CONSENSUS: I-G-[GA]-G-M-[LF]-[SA]-x-P-x(3)-[SA]-G-x(2)-F.

NAME: Xanthine/uracil permeases family signature.
 CONSENSUS: [LIVM]-P-x-[PASIF]-V-[LIVM]-G-G-x(4)-[LIVM]-[FY]-[GSA]-x-[LIVM]-x(3)-G.

NAME: Anion exchangers family signature 1.
 CONSENSUS: F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-R-Y.

NAME: Anion exchangers family signature 2.
 CONSENSUS: [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L.

NAME: MIP family signature.
 CONSENSUS: [HNQA]-x-N-P-[STA]-[LIVMF]-[ST]-[LIVMF]-[GSTAFY].

NAME: General diffusion Gram-negative porins signature.

CONSENSUS: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]-V.

NAME: OmpA-like domain.

CONSENSUS: [LIVMA]-x-[GT]-x-[TA]-[DA]-x(2)-[DG]-[GSTP]-x(2)-[LFYDE]-[NQS]-x(2)-

CONSENSUS: [LJ]-[SG]-[QE]-[KRQE]-R-A-x(2)-[LV]-x(3)-[LIVMF]-x(4,5)-[LIVM]-x(4)-

CONSENSUS: [LIVM]-x(3)-[SG]-x-G.

NAME: Eukaryotic mitochondrial porin signature.

CONSENSUS: [YH]-x(2)-D-[SPA]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]-[DNSTA]-[DNS]-x(4)-

CONSENSUS: [GSTAN]-[LIVMA]-x-[LIVMY].

NAME: Insulin-like growth factor binding proteins signature.

CONSENSUS: G-C-[GS]-C-C-x(2)-C-A-x(6)-C.

NAME: GPR1/FUN34/yaaH family signature.

CONSENSUS: N-P-[AV]-P-[LF]-G-L-x-[GSA]-F.

NAME: GNS1/SUR4 family signature.

CONSENSUS: L-x-F-L-H-x-Y-H-H.

NAME: 43 Kd postsynaptic protein signature.

CONSENSUS: G-Q-D-Q-T-K-Q-Q-I.

NAME: Actins signature 1.

CONSENSUS: [FY]-[LIV]-G-[DE]-E-A-Q-x-[RKQ](2)-G.

NAME: Actins signature 2.

CONSENSUS: W-[IV]-[STA]-[RK]-x-[DE]-Y-[DNE]-[DE].

NAME: Actins and actin-related proteins signature.

CONSENSUS: [LM]-[LIVM]-T-E-[GAPQ]-x-[LIVMFYWHQ]-N-[PSTAQ]-x(2)-N-[KR].

NAME: Annexins repeated domain signature.

CONSENSUS: [TG]-[STV]-x(8)-[LIVMF]-x(2)-R-x(3)-[DEQNH]-x(7)-[IFY]-x(7)-[LIVMF]-

CONSENSUS: x(3)-[LIVMF]-x(11)-[LIVMFA]-x(2)-[LIVMF].

NAME: Caveolins signature.

CONSENSUS: F-E-D-V-I-A-E-P.

NAME: Clathrin light chain signature 1.

CONSENSUS: F-L-A-Q-Q-E-S.

NAME: Clathrin light chain signature 2.

CONSENSUS: [KR]-D-x-S-[KR]-[LIVM]-[KR]-x-[LIVM](3)-x-L-K.

NAME: Clusterin signature 1.

CONSENSUS: C-K-P-C-L-K-x-T-C.

NAME: Clusterin signature 2.

CONSENSUS: C-L-[RK]-M-[RK]-x-[EQ]-C-[ED]-K-C.

NAME: Connexins signature 1.

CONSENSUS: C-[DN]-T-x-Q-P-G-C-x(2)-V-C-Y-D.

NAME: Connexins signature 2.

CONSENSUS: C-x(3,4)-P-C-x(3)-[LIVM]-[DEN]-C-[FY]-[LIVM]-[SA]-[KR]-P.

NAME: Crystallins beta and gamma 'Greek key' motif signature.

CONSENSUS: [LIVMFYWA]-x-[DEHRKSTP]-[FY]-[DEQHKY]-x(3)-[FY]-x-G-x(4)-[LIVMFCST].

NAME: Dynamin family signature.

CONSENSUS: L-P-[RK]-G-[STN]-[GN]-[LIVM]-V-T-R.

NAME: Dynein light chain type 1 signature.

CONSENSUS: H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E.

NAME: FusZ protein signature 1.

CONSENSUS: N-[ST]-D-x-Q-x-L-x(16,18)-G-x-G-[ATV]-G-[GSAN]-x-P-x(2)-G.

NAME: FusZ protein signature 2.

CONSENSUS: [DNHKKR]-[LIVMF]-x-[LIVMF](2)-[VSTAC]-[STAC]-G-x-G-[GK]-G-T-G-[ST]-G-

CONSENSUS: [GSAR]-[STA]-P-[LIVMFT]-[LIVMF]-[SGAV].

NAME: Fungal hydrophobins signature.
 CONSENSUS: [GN]-[DNQPSA]-x-C-[GSTANK]-[GSTADNQ]-[STNQI]-[PTIV]-x-C-C-[DENQKPST].

NAME: Intermediate filaments signature.
 CONSENSUS: [IV]-x-[TACI]-Y-[RKH]-x-[LM]-L-[DE].

NAME: Involucrin signature.
 CONSENSUS: <M-S-[QH]-Q-x-T-[LV]-P-V-T-[LV].

NAME: Kinesin motor domain signature.
 CONSENSUS: [GSA]-[KRHPSTQVM]-[LIVMF]-x-[LIVMF]-[IVC]-D-L-[AH]-G-[SAN]-E.

NAME: Kinesin motor domain profile.

NAME: Kinesin light chain repeat.
 CONSENSUS: [DEQR]-A-L-x(3)-[GEQ]-x(3)-G-x-[DNS]-x-P-x-V-A-x(3)-N-x-L-[AS]-
 CONSENSUS: x(5)-[QR]-x-[KR]-[FY]-x(2)-[AV]-x(4)-[HKNQ].

NAME: Myelin basic protein signature.
 CONSENSUS: V-V-H-F-F-K-N.

NAME: Myelin P0 protein signature.
 CONSENSUS: S-[KR]-S-x-K-[AG]-x-[SA]-E-K-K-[STA]-K.

NAME: Myelin proteolipid protein signature 1.
 CONSENSUS: G-[MV]-A-L-F-C-G-C-G-H.

NAME: Myelin proteolipid protein signature 2.
 CONSENSUS: C-x-[ST]-x-[DE]-x(3)-[ST]-[FY]-x-L-[FY]-I-x(4)-G-A.

NAME: Neuromodulin (GAP-43) signature 1.
 CONSENSUS: <M-L-C-C-[LIVM]-R-R.

NAME: Neuromodulin (GAP-43) signature 2.
 CONSENSUS: S-F-R-G-H-I-x-R-K-K-[LIVM].

NAME: Osteopontin signature.
 CONSENSUS: [KQ]-x-[TA]-x(2)-[GA]-S-S-E-E-K.

NAME: Peripherin / rom-1 signature.
 CONSENSUS: D-[GS]-V-P-F-[ST]-C-C-N-P-x-S-P-R-P-C.

NAME: Profilin signature.
 CONSENSUS: <x(0,1)-[STA]-x(0,1)-W-[DENQH]-x-[YI]-x-[DEQ].

NAME: Surfactant associated polypeptide SP-C palmitoylation sites.
 CONSENSUS: I-P-C-C-P-V.

NAME: Synapsins signature 1.
 CONSENSUS: L-R-R-R-L-S-D-S.

NAME: Synapsins signature 2.
 CONSENSUS: G-H-A-H-S-G-M-G-K-V-K.

NAME: Synaptobrevin signature.
 CONSENSUS: N-[LIVM]-[DENS]-[KL]-V-x-[DEQ]-R-x(2)-[KR]-[LIVM]-[STDE]-x-[LIVM]-x-[DE]-
 CONSENSUS: [KR]-[TA]-[DE].

NAME: Synaptophysin / synaptoporin signature.
 CONSENSUS: L-S-V-[DE]-C-x-N-K-T.

NAME: Tropomyosins signature.
 CONSENSUS: L-K-E-A-E-x-R-A-E.

NAME: Tubulin subunits alpha, beta, and gamma signature.
 CONSENSUS: [SAG]-G-G-T-G-[SA]-G.

NAME: Tubulin-beta mRNA autoregulation signal.
 CONSENSUS: <M-R-[DE]-[IL].

NAME: Tau and MAP proteins tubulin-binding domain signature.
 CONSENSUS: G-S-x(2)-N-x(2)-H-x-[PA]-[AG]-G(2).

NAME: Neuraxin and MAP1B proteins repeated region signature.

CONSENSUS: [STAGDN]-Y-x-Y-E-x(2)-[DE]-[KR]-[STAGCI].
 NAME: F-actin capping protein alpha subunit signature 1.
 CONSENSUS: V-H-[FY](2)-E-D-G-N-V.
 NAME: F-actin capping protein alpha subunit signature 2.
 CONSENSUS: F-K-[AE]-L-R-R-x-L-P.
 NAME: F-actin capping protein beta subunit signature.
 CONSENSUS: C-D-Y-N-R-D.
 NAME: Vinculin family talin-binding region signature.
 CONSENSUS: [KR]-x-[LIVMF]-x(3)-[LIVMA]-x(2)-[LIVM]-x(6)-R-Q-Q-E-L.
 NAME: Vinculin repeated domain signature.
 CONSENSUS: [LIVM]-x-[QA]-A-x(2)-W-[IL]-x-[DN]-P.
 NAME: Amyloidogenic glycoprotein extracellular domain signature.
 CONSENSUS: G-[VT]-E-[FY]-V-C-C-P.
 NAME: Amyloidogenic glycoprotein intracellular domain signature.
 CONSENSUS: G-Y-E-N-P-T-Y-[KR].
 NAME: Cadherins extracellular repeated domain signature.
 CONSENSUS: [LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P.
 NAME: Insect cuticle proteins signature.
 CONSENSUS: G-x(7)-[DEN]-G-x(6)-Y-x-A-[DNG]-x(2,3)-G-[FY]-x-[AP].
 NAME: Gas vesicles protein GVPa signature 1.
 CONSENSUS: [LIVM]-x-[DE]-[LIVMFYT]-[LIVM]-[DE]-x-[LIVM](2)-[DKR](2)-G-x-[LIVM](2).
 NAME: Gas vesicles protein GVPa signature 2.
 CONSENSUS: R-[LIVA](3)-A-[GS]-[LIVMFY]-x-T-x(3)-Y-[AG].
 NAME: Gas vesicles protein GVPc repeated domain signature.
 CONSENSUS: F-L-x(2)-T-x(3)-R-x(3)-A-x(2)-Q-x(3)-L-x(2)-F.
 NAME: Bacterial microcompartments proteins signature.
 CONSENSUS: D-x(0,1)-M-x-K-[SAG](2)-x-[IV]-x-[LIVM]-[LIVMA]-[GCS]-x(4)-[GD]-[SGPD]-[GA].
 NAME: Flagella basal body rod proteins signature.
 CONSENSUS: [GTARYQ]-x(9)-[LIVMYSTA](2)-[GSTA]-[STADEN]-N-[LIVM]-[SAN]-N-x-[SADNFR]-[STV].
 NAME: Flagella transport protein fliP family signature 1.
 CONSENSUS: [PA]-A-[FY]-x-[LIVT]-[STH]-[EQ]-[LI]-x(2)-[GA]-F-[KREQ]-[IM]-G-[LIF].
 NAME: Flagella transport protein fliP family signature 2.
 CONSENSUS: P-[LIVMF]-K-[LIVMF](5)-x-[LIVMA]-[DNKS]-G-W.
 NAME: Plant viruses icosahedral capsid proteins 'S' region signature.
 CONSENSUS: [FYW]-x-[PSTA]-x(7)-G-x-[LIVM]-x-[LIVM]-x-[FYWI]-x(2)-D-x(5)-P.
 NAME: Potexviruses and carlaviruses coat protein signature.
 CONSENSUS: [RK]-[FYW]-A-[GAP]-F-D-x-F-x(2)-[LV]-x(3)-[GAST](2).
 NAME: Neurotransmitter-gated ion-channels signature.
 CONSENSUS: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C.
 NAME: ATP P2X receptors signature.
 CONSENSUS: G-G-x-[LIVM]-G-[LIVM]-x-[IV]-x-W-x-C-[DN]-L-D-x(5)-C-x-P-x-Y-x-F.
 NAME: G-protein coupled receptors signature.
 CONSENSUS: [GSTALIVMFYWC]-[GSTANCPDE]-[EDPKRH]-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM].
 NAME: G-protein coupled receptors family 2 signature 1.
 CONSENSUS: C-x(3)-[FYWLIV]-D-x(3,4)-C-[FW]-x(2)-[STAGV]-x(8,9)-C-[PF].
 NAME: G-protein coupled receptors family 2 signature 2.
 CONSENSUS: Q-G-[LMFCA]-[LIVMFT]-[LIV]-x-[LIVFST]-[LIF]-[VFYH]-C-[LFY]-x-N-x(2)-V.

NAME: G-protein coupled receptors family 3 signature 1.
 CONSENSUS: [LV]-x-N-[LIVM](2)-x-L-F-x-I-[PA]-Q-[LIVM]-[STA]-x-[STA](3)-[STAN].

NAME: G-protein coupled receptors family 3 signature 2.
 CONSENSUS: C-C-[FYW]-x-C-x(2)-C-x(4)-[FYW]-x(2,4)-[DN]-x(2)-[STAH]-C-x(2)-C.

NAME: G-protein coupled receptors family 3 signature 3.
 CONSENSUS: F-N-E-[STA]-K-x-I-[STAG]-F-[ST]-M.

NAME: Visual pigments (opsins) retinal binding site.
 CONSENSUS: [LIVMWAC]-[PGAC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-[STACP]-x(2)-[DENF]-
 CONSENSUS: [AP]-x(2)-[IY].

NAME: Bacterial rhodopsins signature 1.
 CONSENSUS: R-Y-x-[DT]-W-x-[LIVMF]-[ST]-T-P-[LIVM](3).

NAME: Bacterial rhodopsins retinal binding site.
 CONSENSUS: [FYTV]-x-[FYVG]-[LIVM]-D-[LIVMF]-x-[STA]-K-x(2)-[FY].

NAME: Receptor tyrosine kinase class II signature.
 CONSENSUS: [DN]-[LIV]-Y-x(3)-Y-Y-R.

NAME: Receptor tyrosine kinase class III signature.
 CONSENSUS: G-x-H-x-N-[LIVM]-V-N-L-L-G-A-C-T.

NAME: Receptor tyrosine kinase class V signature 1.
 CONSENSUS: F-x-[DN]-x-[GAW]-[GA]-C-[LIVM]-[SA]-[LIVM](2)-[SA]-[LV]-[KRHQ]-[LIVA]-
 CONSENSUS: x(3)-[KR]-C-[PSAW].

NAME: Receptor tyrosine kinase class V signature 2.
 CONSENSUS: C-x(2)-[DE]-G-[DEQ]-W-x(2,3)-[PAQ]-[LIVMT]-[GT]-x-C-x-C-x(2)-G-[HFY]-
 CONSENSUS: [EQ].

NAME: Growth factor and cytokines receptors family signature 1.
 CONSENSUS: C-[LVFYR]-x(7,8)-[STIVDN]-C-x-W.

NAME: Growth factor and cytokines receptors family signature 2.
 CONSENSUS: [STGL]-x-W-[SG]-x-W-S.

NAME: TNFR/NGFR family cysteine-rich region signature.
 CONSENSUS: C-x(4,6)-[FYH]-x(5,10)-C-x(0,2)-C-x(2,3)-C-x(7,11)-C-x(4,6)-[DNEQSKP]-
 CONSENSUS: x(2)-C.

NAME: TNFR/NGFR family cysteine-rich region domain.

NAME: Integrins alpha chain signature.
 CONSENSUS: [FYWS]-[RK]-x-G-F-F-x-R.

NAME: Integrins beta chain cysteine-rich domain signature.
 CONSENSUS: C-x-[GNQ]-x(1,3)-G-x-C-x-C-x(2)-C-x-C.

NAME: Natriuretic peptides receptors signature.
 CONSENSUS: G-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W.

NAME: Photosynthetic reaction center proteins signature.
 CONSENSUS: [NH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2).

NAME: Antenna complexes alpha subunits signature.
 CONSENSUS: [LIVFAG]-x-[GASV]-[LIVFA]-x-[IV]-H-x(3)-[LIVM]-[GSTAE]-[STANH]-x(1,3)-
 CONSENSUS: [STN]-W-[LIVMFYW].

NAME: Antenna complexes beta subunits signature.
 CONSENSUS: [EQ]-x(4)-H-x(5)-[GSTA]-x(3)-[FY]-x(3)-[AG]-x(2)-[AV]-H-x(7)-P.

NAME: Photosystem I psaA and psaB proteins signature.
 CONSENSUS: C-D-G-P-G-R-G-G-T-C.

NAME: Photosystem I psaG and psaK proteins signature.
 CONSENSUS: G-F-x-[LIVM]-x-[DEA]-x(2)-[GA]-x-[GTA]-[SA]-x-G-H-x-[LIVM]-[GA].

NAME: Phytochrome chromophore attachment site signature.
 CONSENSUS: [RGS]-[GSA]-[PV]-H-x-C-H-x(2)-Y.

NAME: Phytochrome chromophore attachment site domain profile.

NAME: Speract receptor repeated domain signature.

CONSENSUS: G-x(5)-G-x(2)-E-x(6)-W-G-x(2)-C-x(3)-[FYW]-x(8)-C-x(3)-G.

NAME: TonB-dependent receptor proteins signature 1.

CONSENSUS: < x(10,115)-[DENF]-[ST]-[LIVMF]-[LIVSTEQ]-V-x-[AGP]-[STANEQPK].

NAME: TonB-dependent receptor proteins signature 2.

CONSENSUS: [LYGSTANE]-x(3)-[GSTAENQ]-x-[PGE]-R-x-[LIVFYWA]-x-[LIVMFTA]-[STAGNQ]-

CONSENSUS: [LIVMFYGT]-x-[LIVMFYWGTDQ]-x-F>.

NAME: Transmembrane 4 family signature.

CONSENSUS: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF](2)-G-C-x-[GA]-[STA]-x(2)-[EG]-x(2)-

CONSENSUS: [CWN]-[LIVM](2).

NAME: Bacterial chemotaxis sensory transducers signature.

CONSENSUS: R-T-E-[EQ]-Q-x(2)-[SA]-[LIVM]-x-[EQ]-T-A-A-S-M-E-Q-L-T-A-T-V.

NAME: ER lumen protein retaining receptor signature 1.

CONSENSUS: G-I-S-x-[KR]-x-Q-x-L-[FY]-x-[LIV](2)-F-x(2)-R-Y.

NAME: ER lumen protein retaining receptor signature 2.

CONSENSUS: L-E-[SA]-V-A-I-[LM]-P-Q-L.

NAME: Ephrins signature.

CONSENSUS: [KRQ]-[LF]-[CST]-x-K-[IF]-Q-x-[FY]-[ST]-[PA]-x(3)-G-x-E-F-x(5)-[FY](2)-

CONSENSUS: x(2)-[SA].

NAME: Granulins signature.

CONSENSUS: C-x-D-x(2)-H-C-C-P-x(4)-C.

NAME: HBGF/FGF family signature.

CONSENSUS: G-x-L-x-[STAGP]-x(6,7)-[DE]-C-x-[FM]-x-E-x(6)-Y.

NAME: PTN/MK heparin-binding protein family signature 1.

CONSENSUS: S-[DE]-C-x-[DE]-W-x-W-x(2)-C-x-P-x-[SN]-x-D-C-G-[LIVMA]-G-x-R-E-G.

NAME: PTN/MK heparin-binding protein family signature 2.

CONSENSUS: C-[KR]-[LIVM]-P-C-N-W-K-K-x-F-G-A-[DE]-C-K-Y-x-F-[EQ]-x-W-G-x-C.

NAME: Nerve growth factor family signature.

CONSENSUS: G-C-[KR]-G-[LIV]-[DE]-x(3)-[YW]-x-S-x-C.

NAME: Platelet-derived growth factor (PDGF) family signature.

CONSENSUS: P-[PS]-C-V-x(3)-R-C-[GSTA]-G-C-C.

NAME: Small cytokines (intercrine/chemokine) C-x-C subfamily signature.

CONSENSUS: C-x-C-[LIVM]-x(5,6)-[LIVMFY]-x(2)-[RKSEQ]-x-[LIVM]-x(2)-[LIVM]-x(5)-

CONSENSUS: [SAG]-x(2)-C-x(3)-[EQ]-[LIVM](2)-x(9,10)-C-L-[DN].

NAME: Small cytokines (intercrine/chemokine) C-C subfamily signature.

CONSENSUS: C-C-[LIFYT]-x(5,6)-[LI]-x(4)-[LIVMF]-x(2)-[FYW]-x(6,8)-C-x(3,4)-[SAG]-

CONSENSUS: [LIVM](2)-[FL]-x(8)-C-[STA].

NAME: TGF-beta family signature.

CONSENSUS: [LIVM]-x(2)-P-x(2)-[FY]-x(4)-C-x-G-x-C.

NAME: TNF family signature.

CONSENSUS: [LV]-x-[LIVM]-x(3)-G-[LIVMF]-Y-[LIVMFY](2)-x(2)-[QEKHL]-[LIVMGT]-x-

CONSENSUS: [LIVMFY].

NAME: TNF family profile.

NAME: Wnt-1 family signature.

CONSENSUS: C-K-C-H-G-[LIVMT]-S-G-x-C.

NAME: Interferon alpha, beta and delta family signature.

CONSENSUS: [FYH]-[FY]-x-[GNRC]-[LIVM]-x(2)-[FY]-L-x(7)-[CY]-A-W.

NAME: Granulocyte-macrophage colony-stimulating factor signature.

CONSENSUS: C-P-[LP]-T-x-E-[ST]-x-C.

NAME: Interleukin-1 signature.

CONSENSUS: [FC]-x-S-[ASLV]-x(2)-P-x(2)-[FYLIV]-[LI]-[SCA]-T-x(7)-[LIVM].

NAME: Interleukin-2 signature.

CONSENSUS: T-E-[LF]-x(2)-L-x-C-L-x(2)-E-L.

NAME: Interleukins -4 and -13 signature.

CONSENSUS: L-x-E-[LIVM](2)-x(4,5)-[LIVM]-[TL]-x(5,7)-C-x(4)-[IVA]-x-[DNS]-[LIVMA].

NAME: Interleukin-6 / G-CSF / MGF signature.

CONSENSUS: C-x(9)-C-x(6)-G-L-x(2)-[FY]-x(3)-L.

NAME: Interleukin-7 and -9 signature.

CONSENSUS: N-x-[LAP]-[SCT]-F-L-K-x-L-L.

NAME: Interleukin-10 family signature.

CONSENSUS: [GS]-C-x(2)-[LV]-x(2)-[LIVM](2)-x-F-Y-L-x(2)-V.

NAME: LIF / OSM family signature.

CONSENSUS: [PST]-x(4)-F-[NQ]-x-K-x(3)-C-x-[LF]-L-x(2)-Y-[HK].

NAME: Macrophage migration inhibitory factor family signature.

CONSENSUS: [DE]-P-C-A-x(3)-[LIVM]-x-S-I-G-x-[LIVM]-G.

NAME: Adipokinetic hormone family signature.

CONSENSUS: Q-[LV]-[NT]-[FY]-[ST]-x(2)-W.

NAME: Bombesin-like peptides family signature.

CONSENSUS: W-A-x-G-[SH]-[LF]-M.

NAME: Calcitonin / CGRP / IAPP family signature.

CONSENSUS: C-[SAGDN]-[STN]-x(0,1)-[SA]-T-C-[VMA]-x(3)-[LYF]-x(3)-[LYF].

NAME: Corticotropin-releasing factor family signature.

CONSENSUS: [PQ]-x-[LIVM]-S-[LIVM]-x(2)-[PST]-[LIVMF]-x-[LIVM]-L-R-x(2)-[LIVM].

NAME: Crustacean CHH/MIH/GIH neurohormones family signature.

CONSENSUS: C-[DENK]-D-C-x-N-[LIV]-[FY]-R-x(7)-C-[KR]-x(2)-C.

NAME: Erythropoietin / thrombopoietin signature.

CONSENSUS: P-x(4)-C-D-x-R-[LIVM](2)-x-[KR]-x(14)-C.

NAME: Granins signature 1.

CONSENSUS: [DE]-[SN]-L-[SAN]-x(2)-[DE]-x-E-L.

NAME: Granins signature 2.

CONSENSUS: C-[LIVM](2)-E-[LIVM](2)-S-[DN]-[STA]-L-x-K-x-S-x(3)-[LIVM]-[STA]-x-E-C.

NAME: Galanin signature.

CONSENSUS: G-W-T-L-N-S-A-G-Y-L-L-G-P-H.

NAME: Gastrin / cholecystokinin family signature.

CONSENSUS: Y-x(0,1)-[GD]-[WH]-M-[DR]-F.

NAME: Glucagon / GIP / secretin / VIP family signature.

CONSENSUS: [YH]-[STAIVGD]-[DEQ]-[AGF]-[LIVMSTE]-[FYLR]-x-[DENSTAK]-[DENSTA]-
[LIVMFYG]-x(9)-[KREQL]-[KRDENQL]-[LVFYWG]-[LIVQ].

NAME: Glycoprotein hormones alpha chain signature 1.

CONSENSUS: C-x-G-C-C-[FY]-S-R-A-[FY]-P-T-P.

NAME: Glycoprotein hormones alpha chain signature 2.

CONSENSUS: N-H-T-x-C-x-C-x-T-C-x(2)-H-K.

NAME: Glycoprotein hormones beta chain signature 1.

CONSENSUS: C-[STAGM]-G-[HFYL]-C-x-[ST].

NAME: Glycoprotein hormones beta chain signature 2.

CONSENSUS: [PA]-V-A-x(2)-C-x-C-x(2)-C-x(4)-[STD]-[DEY]-C-x(6,8)-[PGSTAVM]-x(2)-C.

NAME: Gonadotropin-releasing hormones signature.

CONSENSUS: Q-H-[FYW]-S-x(4)-P-G.

NAME: Insulin family signature.

CONSENSUS: C-C-[P]-x(2)-C-[STDNEKPT]-x(3)-[LIVMFS]-x(3)-C.

NAME: Natriuretic peptides signature.

CONSENSUS: C-F-G-x(3)-D-R-I-x(3)-S-x(2)-G-C.

NAME: Neurohypophysial hormones signature.

CONSENSUS: C-[LIFY](2)-x-N-[CS]-P-x-G.

NAME: Neuromedin U signature.

CONSENSUS: F-[LIVMF]-F-R-P-R-N.

NAME: Endogenous opioids neuropeptides precursors signature.

CONSENSUS: C-x(3)-C-x(2)-C-x(2)-[KRH]-x(6,7)-[LIF]-[DN]-x(3)-C-x-[LIVM]-[EQ]-C-

CONSENSUS: [EQ]-x(8)-W-x(2)-C.

NAME: Pancreatic hormone family signature.

CONSENSUS: [FY]-x(3)-[LIVM]-x(2)-Y-x(3)-[LIVMFY]-x-R-x-R-[YF].

NAME: Parathyroid hormone family signature.

CONSENSUS: V-S-E-x-Q-x(2)-H-x(2)-G.

NAME: Pyrokinins signature.

CONSENSUS: F-[GSTV]-P-R-L-[G>].

NAME: Somatotropin, prolactin and related hormones signature 1.

CONSENSUS: C-x-[ST]-x(2)-[LIVMFY]-x-[LIVMSTA]-P-x(5)-[TALIV]-x(7)-[LIVMFY]-x(6)-

CONSENSUS: [LIVMFY]-x(2)-[STA]-W.

NAME: Somatotropin, prolactin and related hormones signature 2.

CONSENSUS: C-[LIVMFY]-x(2)-D-[LIVMFYSTA]-x(5)-[LIVMFY]-x(2)-[LIVMFYT]-x(2)-C.

NAME: Tachykinin family signature.

CONSENSUS: F-[IVFY]-G-[LM]-M-[G>].

NAME: Thymosin beta-4 family signature.

CONSENSUS: K-L-K-K-T-E-T-Q-E-K-N.

NAME: Urotensin II signature.

CONSENSUS: C-F-W-K-Y-C.

NAME: Cecropin family signature.

CONSENSUS: W-x(0,2)-[KDN]-x(2)-K-[KRE]-[LI]-E-[RKN].

NAME: Mammalian defensins signature.

CONSENSUS: C-x-C-x(3,5)-C-x(7)-G-x-C-x(9)-C-C.

NAME: Arthropod defensins signature.

CONSENSUS: C-x(2,3)-[HN]-C-x(3,4)-[GR]-x(2)-G-G-x-C-x(4,7)-C-x-C.

NAME: Cathelicidins signature 1.

CONSENSUS: Y-x-[ED]-x-V-x-[RQ]-A-[LIVMA]-[DQG]-x-[LIVMFY]-N-[EQ].

NAME: Cathelicidins signature 2.

CONSENSUS: F-x-[LIVM]-K-E-T-x-C-x(10)-C-x-F-[KR]-[KE].

NAME: Endothelin family signature.

CONSENSUS: C-x-C-x(4)-D-x(2)-C-x(2)-[FY]-C.

NAME: Plant thionins signature.

CONSENSUS: C-C-x(5)-R-x(2)-[FY]-x(2)-C.

NAME: Gamma-thionins family signature.

CONSENSUS: [KR]-x-C-x(3)-[SV]-x(2)-[FYWH]-x-[GF]-x-C-x(5)-C-x(3)-C.

NAME: Snake toxins signature.

CONSENSUS: G-C-x(1,3)-C-P-x(8,10)-C-C-x(2)-[PDEN].

NAME: Myotoxins signature.

CONSENSUS: K-x-C-H-x-K-x(2)-H-C-x(2)-K-x(3)-C-x(8)-K-x(2)-C-x(2)-[RK]-x-K-C-C-K-K.

NAME: Scorpion short toxins signature.

CONSENSUS: C-x(3)-C-x(6,9)-[GAS]-K-C-[IMQT]-x(3)-C-x-C.

NAME: Heat-stable enterotoxins signature.

CONSENSUS: C-C-x(2)-C-C-x-P-A-C-x-G-C.

NAME: Aerolysin type toxins signature.

CONSENSUS: [KT]-x(2)-N-W-x(2)-T-[DN]-T.

NAME: Shiga/ricin ribosomal inactivating toxins active site signature.

CONSENSUS: [LIVMA]-x-[LIVMSTA](2)-x-E-[SAGV]-[STAL]-R-[FY]-[RKNQS]-x-[LIVM]-[EQS]-

CONSENSUS: x(2)-[LIVMF].

NAME: Channel forming colicins signature.

CONSENSUS: T-x(2)-W-x-P-[LIVMFY](3)-x(2)-E.

NAME: Hok/gef family cell toxic proteins signature.

CONSENSUS: [LIVMA](4)-C-[LIVMFA]-T-[LIVMA](2)-x(4)-[LIVM]-x-[RG]-x(2)-L-[CY].

NAME: Staphylococcal enterotoxin/Streptococcal pyrogenic exotoxin signature 1.

CONSENSUS: Y-G-G-[LIV]-T-x(4)-N.

NAME: Staphylococcal enterotoxin/Streptococcal pyrogenic exotoxin signature 2.

CONSENSUS: K-x(2)-[LIV]-x(4)-[LIV]-D-x(3)-R-x(2)-L-x(5)-[LIV]-Y.

NAME: Thiol-activated cytolysins signature.

CONSENSUS: [RK]-E-C-T-G-L-x-W-E-W-W-[RK].

NAME: Membrane attack complex components / perforin signature.

CONSENSUS: Y-x(6)-[FY]-G-T-H-[FY].

NAME: Pancreatic trypsin inhibitor (Kunitz) family signature.

CONSENSUS: F-x(3)-G-C-x(6)-[FY]-x(5)-C.

NAME: Bowman-Birk serine protease inhibitors family signature.

CONSENSUS: C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDKS]-[DEKRHSTA]-C.

NAME: Kazal serine protease inhibitors family signature.

CONSENSUS: C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C.

NAME: Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature.

CONSENSUS: [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM].

NAME: Serpins signature.

CONSENSUS: [LIVMFY]-x-[LIVMFYAC]-[DNQ]-[RKHQSS]-[PST]-F-[LIVMFY]-[LIVMFYC]-x-

CONSENSUS: [LIVMFAH].

NAME: Potato inhibitor I family signature.

CONSENSUS: [FYW]-P-[EQH]-[LIV](2)-G-x(2)-[STAGV]-x(2)-A.

NAME: Squash family of serine protease inhibitors signature.

CONSENSUS: C-P-x(5)-C-x(2)-D-x-D-C-x(3)-C-x-C.

NAME: Streptomyces subtilisin-type inhibitors signature.

CONSENSUS: C-x-P-x(2,3)-G-x-H-P-x(4)-A-C-[ATD]-x-L.

NAME: Cysteine proteases inhibitors signature.

CONSENSUS: [GSTEQKRV]-Q-[LIVT]-[VAF]-[SAGQ]-G-x-[LIVMNK]-x(2)-[LIVMFY]-x-[LIVMFYA]-

CONSENSUS: [DENQKRHSIV].

NAME: Tissue inhibitors of metalloproteinases signature.

CONSENSUS: C-x-C-x-P-x-H-P-Q-x-A-F-C.

NAME: Cereal trypsin/alpha-amylase inhibitors family signature.

CONSENSUS: C-x(4)-[SAGD]-x(4)-[SPAL]-[LF]-x(2)-C-[RH]-x-[LIVMFY](2)-x(3,4)-C.

NAME: Alpha-2-macroglobulin family thiolester region signature.

CONSENSUS: [PG]-x-[GS]-C-[GA]-E-[EQ]-x-[LIVM].

NAME: Disintegrins signature.

CONSENSUS: C-x(2)-G-x-C-C-x-[NQRS]-C-x-[FM]-x(6)-C-[RK].

NAME: Lambdoid phages regulatory protein CIII signature.

CONSENSUS: E-S-x-L-x-R-x(2)-[KR]-x-L-x(4)-[KR](2)-x(2)-[DE]-x-L.

NAME: Chaperonins cpn60 signature.

CONSENSUS: A-[AS]-x-[DEQ]-E-x(4)-G-G-[GA].

NAME: Chaperonins cpn10 signature.

CONSENSUS: [LIVMFY]-x-P-[ILT]-x-[DEN]-[KR]-[LIVMFA](3)-[KREQ]-x(8,9)-[SG]-x-

CONSENSUS: [LIVMFY](3).

NAME: Chaperonins TCP-1 signature 1.

CONSENSUS: [RKEL]-[ST]-x-[LMFY]-G-P-x-[GSA]-x-x-K-[LIVMF](2).

NAME: Chaperonins TCP-1 signature 2.

CONSENSUS: [LIVM]-[TS]-[NK]-D-[GA]-[AVNHK]-[TAV]-[LIVM](2)-x(2)-[LIVM]-x-[LIVM]-x-

CONSENSUS: [SNH]-[POH].

NAME: Chaperonins TCP-1 signature 3.

CONSENSUS: Q-[DEK]-x-x-[LIVMGTA]-[GA]-D-G-T.

NAME: Heat shock hsp20 proteins family profile.

NAME: Heat shock hsp70 proteins family signature 1.

CONSENSUS: [IV]-D-L-G-T-[ST]-x-[SC].

NAME: Heat shock hsp70 proteins family signature 2.

CONSENSUS: [LIVMF]-[LIVMFY]-[DN]-[LIVMFS]-G-[GSH]-[GS]-[AST]-x(3)-[ST]-[LIVM]-

CONSENSUS: [LIVMFC].

NAME: Heat shock hsp70 proteins family signature 3.

CONSENSUS: [LIVMY]-x-[LIVMF]-x-G-G-x-[ST]-x-[LIVM]-P-x-[LIVM]-x-[DEQKRSTA].

NAME: Heat shock hsp90 proteins family signature.

CONSENSUS: Y-x-[NQH]-K-[DE]-[IVA]-F-L-R-[ED].

NAME: Chaperonins clpA/B signature 1.

CONSENSUS: D-[AI]-[SGA]-N-[LIVMF](2)-K-[PT]-x-L-x(2)-G.

NAME: Chaperonins clpA/B signature 2.

CONSENSUS: R-[LIVMFY]-D-x-S-E-[LIVMFY]-x-E-[KRQ]-x-[STA]-x-[STA]-[KR]-[LIVM]-x-G-

CONSENSUS: [STA].

NAME: Nt-dnaJ domain signature.

CONSENSUS: [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)-[KR]-x(2)-[FYI].

NAME: dnaJ domain profile.

NAME: CXXCXGXXG dnaJ domain signature.

CONSENSUS: C-[DEGSTHKKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C-x(2,3)-C-x-G-x-G.

NAME: grpE protein signature.

CONSENSUS: [FL]-[DN]-[PHEA]-x(2)-[HM]-x-A-[LIVMTN]-x(16,20)-G-[FY]-x(3)-[DEG]-x(2)-

CONSENSUS: [LIVM]-[RI]-x-[SA]-x-V-x-[IV].

NAME: Bacterial type II secretion system protein C signature.

CONSENSUS: P-x(6)-F-x(4)-L-x(3)-D-[LIVM]-A-[LIVM]-x-[LIVM]-N-x-[LIVM]-x-L.

NAME: Bacterial type II secretion system protein D signature.

CONSENSUS: [GR]-[DEQKG]-[STVM]-[LIVMA](3)-[GA]-G-[LIVMFY]-x(11)-[LIVM]-P-

CONSENSUS: [LIVMFYWGS]-[LIVMF]-[GSAE]-x-[LIVM]-P-[LIVMFYW](2)-x(2)-[LV]-F.

NAME: Bacterial type II secretion system protein E signature.

CONSENSUS: [LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D.

NAME: Bacterial type II secretion system protein F signature.

CONSENSUS: [KRQ]-[LIVMA]-x(2)-[SAIV]-[LIVM]-x-[TY]-P-x(2)-[LIVM]-x(3)-[STAGV]-x(6)-

CONSENSUS: [LMY]-x(3)-[LIVMF](2)-P.

NAME: Bacterial type II secretion system protein N signature.

CONSENSUS: G-T-L-W-x-G-x(11)-L-x(4)-W.

NAME: Bacterial export FHIPEP family signature.

CONSENSUS: R-[LIVM]-[GSA]-E-V-[GSA]-A-R-F-[STV]-L-D-[GSA]-M-P-G-K-Q-M-[GSA]-I-D-

CONSENSUS: [GSA]-D.

NAME: Protein secA signatures.

CONSENSUS: [IV]-x-[IV]-[SA]-T-[NQ]-M-A-G-R-G-x-D-I-x-L.

NAME: Protein secY signature 1.

CONSENSUS: [GST]-[LIVMF](2)-x-[LIVM]-G-[LIVM]-x-P-[LIVMFY](2)-x-[AS]-[IGSTQ]-

CONSENSUS: [LIVMFAT](3)-Q-[LIVMFA](2).

NAME: Protein secY signature 2.

CONSENSUS: [LIVMFYW](2)-x-[DE]-x-[LIVMF]-[STN]-x(2)-G-[LIVMF]-[GST]-[NST]-G-x-[GST]-
CONSENSUS: [LIVMF](3).

NAME: Protein secE/sec61-gamma signature.

CONSENSUS: [LIVMFY]-x(2)-[DENQGA]-x(4)-[LIVMTA]-x-[KRV]-x(2)-[KW]-P-x(3)-[SEQ]-x(7)-
CONSENSUS: [LIVT]-[LIVGA]-[LIVFGAST].

NAME: Gram-negative pili assembly chaperone signature.

CONSENSUS: [LIVMFY]-[APN]-x-[DNS]-[KREQ]-E-[STR]-[LIVMAR]-x-[FYWT]-x-[NC]-[LIVM]-
CONSENSUS: x(2)-[LIVM]-P-[PAS].

NAME: Fimbrial biogenesis outer membrane usher protein signature.

CONSENSUS: [VL]-[PASQ]-[PAS]-G-[PAD]-[FY]-x-[LI]-[DNQSTAP]-[DNH]-[LIVMFY].

NAME: SRP54-type proteins GTP-binding domain signature.

CONSENSUS: P-[LIVM]-x-[FYL]-[LIVMAT]-[GS]-x-[GS]-[EQ]-x(4)-[LIVMF].

NAME: Cytochrome c oxidase assembly factor COX10/ctaB/cyoE signature.

CONSENSUS: [ED]-x-D-x(2)-M-x-R-T-x(2)-R-x(4)-G.

NAME: Cyclin-dependent kinases regulatory subunits signature 1.

CONSENSUS: Y-S-x-[KR]-Y-x-[DE](2)-x-[FY]-E-Y-R-H-V-x-[LV]-[PT]-[KRP].

NAME: Cyclin-dependent kinases regulatory subunits signature 2.

CONSENSUS: H-x-P-E-x-H-[IV]-L-L-F-[KR].

NAME: Pentaxin family signature.

CONSENSUS: H-x-C-x-[ST]-W-x-[ST].

NAME: Immunoglobulins and major histocompatibility complex proteins signature.

CONSENSUS: [FY]-x-C-x-[VA]-x-H.

NAME: Prion protein signature 1.

CONSENSUS: A-G-A-A-A-A-G-A-V-V-G-G-L-G-G-Y.

NAME: Prion protein signature 2.

CONSENSUS: E-x-[ED]-x-K-[LIVM](2)-x-[KR]-[LIVM](2)-x-[QE]-M-C-x(2)-Q-Y.

NAME: Cyclins signature.

CONSENSUS: R-x(2)-[LIVMSA]-x(2)-[FYWS]-[LIVM]-x(8)-[LIVMFC]-x(4)-[LIVMFYA]-x(2)-
CONSENSUS: [STAGC]-[LIVMFYQ]-x-[LIVMFYC]-[LIVMFY]-D-[RKH]-[LIVMFYW].

NAME: Proliferating cell nuclear antigen signature 1.

CONSENSUS: [GA]-[LIVMF]-x-[LIVMA]-x-[SAV]-[LIVM]-D-x-[NSAE]-[HKR]-[VI]-x-[LY]-
CONSENSUS: [VGA]-x-[LIVM]-x-[LIVM]-x(4)-F.

NAME: Proliferating cell nuclear antigen signature 2.

CONSENSUS: [RKA]-C-[DE]-[RH]-x(3)-[LIVMF]-x(3)-[LIVM]-x-[SGAN]-[LIVMF]-x-K-
CONSENSUS: [LIVMF](2).

NAME: Actin-depolymerizing proteins signature.

CONSENSUS: P-[DE]-x-[SA]-x-[LIVMT]-[KR]-x-[KR]-M-[LIVM]-[YA]-[STA](3)-x(3)-[LIVMF]-
CONSENSUS: [KR].

NAME: BCL2-like apoptosis inhibitors (spans part of BH3, BH1 and BH2).

NAME: Apoptosis regulator, Bcl-2 family BH1 domain signature.

CONSENSUS: [LVME]-[FT]-x-[GSD]-[GL]-x(1,2)-[NS]-[YW]-G-R-[LIV]-[LIVC]-[GAT]-
CONSENSUS: [LIVMF](2)-x-F-[GSAE]-[GSARY].

NAME: Apoptosis regulator, Bcl-2 family BH2 domain signature.

CONSENSUS: W-[LIM]-x(3)-[GR]-G-[WQ]-[DENSAV]-x-[FLGA]-[LIVFTC].

NAME: Apoptosis regulator, Bcl-2 family BH3 domain signature.

CONSENSUS: [LIVAT]-x(3)-L-[KARQ]-x-[IVAL]-G-D-[DESG]-[LIMFV]-[DENS HQ]-[LVSHRQ]-
CONSENSUS: [NSR].

NAME: Apoptosis regulator, Bcl-2 family BH4 domain signature.

CONSENSUS: [DS]-[NT]-R-[AE]-[LI]-V-x-[KD]-[FY]-[LIV]-[GHS]-Y-K-L-[SR]-Q-[RK]-G-
CONSENSUS: [HY]-x-[CW].

NAME: Apoptosis regulator, Bcl-2 family BH4 domain profile.

NAME: Arrestins signature.

CONSENSUS: [FY]-R-Y-G-x-[DE](2)-x-[DE]-[LIVM](2)-G-[LIVM]-x-F-x-[RK]-[DEQ]-[LIVM].

NAME: AAA-protein family signature.

CONSENSUS: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]-D-x-A-[LIFA]-

CONSENSUS: x-R.

NAME: Ubiquitin domain signature.

CONSENSUS: K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]-[LIVMC]-

CONSENSUS: [LIVMFY]-x-G-x(4)-[DE].

NAME: Ubiquitin domain profile.

NAME: ADP-ribosylation factors family signature.

CONSENSUS: [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)-[GSA]-[LIVMF]-x-

CONSENSUS: [WK]-[LIVM].

NAME: GTP-binding nuclear protein ran signature.

CONSENSUS: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.

NAME: SAR1 family signature.

CONSENSUS: R-x-[LIVM]-E-V-F-M-C-S-[LIVM](2)-x-[KRQ]-x-G-Y-x-E-[AG]-[FI]-x-W-[LIVM]-

CONSENSUS: x-Q-Y.

NAME: Band 7 protein family signature.

CONSENSUS: R-x(2)-[LIV]-[SAN]-x(6)-[LIV]-D-x(2)-T-x(2)-W-G-[LIV]-[KRH]-[LIV]-x-

CONSENSUS: [KR]-[LIV]-E-[LIV]-[KR].

NAME: Trp-Asp (WD) repeats signature.

CONSENSUS: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-

CONSENSUS: [LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN].

NAME: G-protein gamma subunit profile.

NAME: Ras GTPase-activating proteins signature.

CONSENSUS: [GSN]-x-[LIVMF]-[FY]-[LIVMFY]-R-[LIVMFY](2)-[GACN]-P-[AV]-[LIV](2)-

CONSENSUS: [SGAN]-P.

NAME: Ras GTPase-activating proteins profile.

NAME: Guanine-nucleotide dissociation stimulators CDC24 family signature.

CONSENSUS: L-x(2)-[LIVMFYW]-L-x(2)-P-[LIVM]-x(2)-[LIVM]-x-[KRS]-x(2)-L-x-[LIVM]-x-

CONSENSUS: [DEQ]-[LIVM]-x(3)-[ST].

NAME: Guanine-nucleotide dissociation stimulators CDC25 family signature.

CONSENSUS: [GAP]-[CT]-V-P-[FY]-x(4)-[LIVMFY]-x-[DN]-[LIVM].

NAME: MARCKS family signature 1.

CONSENSUS: G-Q-E-N-G-H-V-[KR].

NAME: MARCKS family phosphorylation site domain.

CONSENSUS: E-T-P-K(5)-x(0,1)-F-S-F-K-K-x-F-K-L-S-G-x-S-F-K-[KR]-[NS]-[KR]-K-E.

NAME: Stathmin family signature 1.

CONSENSUS: P-[KQ]-[KR](2)-[DE]-x-S-L-[EG]-E.

NAME: Stathmin family signature 2.

CONSENSUS: A-E-K-R-E-H-E-[KR]-E-V.

NAME: GTP-binding elongation factors signature.

CONSENSUS: D-[KRSTGANQFYW]-x(3)-E-[KRAQ]-x-[RKQD]-[GC]-[IVMK]-[ST]-[IV]-x(2)-

CONSENSUS: [GSTACKRNQ].

NAME: Elongation factor 1 beta/beta'/delta chain signature 1.

CONSENSUS: [DE]-[DEG]-[DE](2)-[LIVMF]-D-L-F-G.

NAME: Elongation factor 1 beta/beta'/delta chain signature 2.

CONSENSUS: V-Q-S-x-D-[LIVM]-x-A-[FWM]-[NQ]-K-[LIVM].

NAME: Elongation factor 1 gamma chain profile.

NAME: Elongation factor Ts signature 1.

CONSENSUS: L-R-x(2)-T-[GDQ]-x-[GS]-[LIVMF]-x(0,1)-[DENKAC]-x-K-[KRNEQS]-[AV]-L.

NAME: Elongation factor Ts signature 2.

CONSENSUS: E-[LIVM]-N-[SCV]-[QE]-T-D-F-V-[SA]-[KRN].

NAME: Elongation factor P signature.

CONSENSUS: K-x-A-x(4)-G-x(2)-[LIV]-x-V-P-x(2)-[LIV]-x(2)-G.

NAME: Eukaryotic initiation factor 1A signature.

CONSENSUS: [IM]-x-G-x-[GS]-[KRH]-x(4)-[CL]-x-D-G-x(2)-R-x(2)-[RH]-I-x-G.

NAME: Eukaryotic initiation factor 4E signature.

CONSENSUS: [DE]-[IFY]-x(2)-F-[KR]-x(2)-[LIVM]-x-P-x-W-E-[DV]-x(5)-G-G-[KR]-W.

NAME: Eukaryotic initiation factor 5A hypusine signature.

CONSENSUS: [PT]-G-K-H-G-x-A-K.

NAME: Initiation factor 2 signature.

CONSENSUS: G-x-[LIVM]-x(2)-L-[KR]-[KRHNS]-x-K-x(5)-[LIVM]-x(2)-G-x-[DEN]-C-G.

NAME: Initiation factor 3 signature.

CONSENSUS: [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-[LIVMFYS]-x-[FY]-[DEQT]-x(2)-[KR].

NAME: Translation initiation factor SUII signature.

CONSENSUS: [LIVM]-[EQ]-[LIVM]-Q-G-[DEN]-[KHQ]-[KRV].

NAME: Prokaryotic-type class I peptide chain release factors signature.

CONSENSUS: [AR]-[STA]-x-G-x-G-G-Q-[HNGCS]-V-N-x(3)-[ST]-A-[IV].

NAME: Transcription termination factor nusG signature.

CONSENSUS: [LIVM]-F-G-[KRW]-x-T-P-[IV]-x-[LIVM].

NAME: Calponin family repeat.

CONSENSUS: [LIVM]-x-[LS]-Q-[MAS]-G-[STY]-[NT]-[KRQ]-x(2)-[STN]-Q-x-G-x(3,4)-G.

NAME: CAP protein signature 1.

CONSENSUS: [LIVM](2)-x-R-L-[DE]-x(4)-R-L-E.

NAME: CAP protein signature 2.

CONSENSUS: D-[LIVMFY]-x-E-x-[PA]-x-P-E-Q-[LIVMFY]-K.

NAME: Calreticulin family signature 1.

CONSENSUS: [KRHN]-x-[DEQN]-[DEQNK]-x(3)-C-G-G-[AG]-[FY]-[LIVM]-[KN]-[LIVMFY](2).

NAME: Calreticulin family signature 2.

CONSENSUS: [LIVM](2)-F-G-P-D-x-C-[AG].

NAME: Calreticulin family repeated motif signature.

CONSENSUS: [IV]-x-D-x-[DENST]-x(2)-K-P-[DEH]-D-W-[DEN].

NAME: Calsequestrin signature 1.

CONSENSUS: [EQ]-[DE]-G-L-[DN]-F-P-x-Y-D-G-x-D-R-V.

NAME: Calsequestrin signature 2.

CONSENSUS: [DE]-L-E-D-W-[LIVM]-E-D-V-L-x-G-x-[LIVM]-N-T-E-D-D-D.

NAME: S-100/ICaBP type calcium binding protein signature.

CONSENSUS: [LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x-[ES]-[FYVC]-x(2)-

CONSENSUS: [LIVMFS]-[LIVMF].

NAME: Hemolysin-type calcium-binding region signature.

CONSENSUS: D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D.

NAME: HlyD family secretion proteins signature.

CONSENSUS: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-

CONSENSUS: [LIVMFYW](2)-x-[LIVMFYW](3).

NAME: P-II protein urydlation site.

CONSENSUS: Y-[KR]-G-[AS]-[AE]-Y.

NAME: P-II protein C-terminal region signature.

CONSENSUS: [ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM].

NAME: 14-3-3 proteins signature 1.

CONSENSUS: R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA].

NAME: 14-3-3 proteins signature 2.

CONSENSUS: Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD].

NAME: ATP1G1 / PLM / MAT8 family signature.

CONSENSUS: [DNS]-x-F-x-Y-D-x(2)-[ST]-[LIVM]-[RQ]-x(2)-G.

NAME: BTG1 family signature 1.

CONSENSUS: Y-x(2)-[HP]-W-[FY]-[AP]-E-x-P-x-K-G-x-[GA]-[FY]-R-C-[IV]-[RH]-[IV].

NAME: BTG1 family signature 2.

CONSENSUS: [LV]-P-x-[DE]-[LM]-[ST]-[LIVM]-W-[IV]-D-P-x-E-V-[SC]-x-[RQ]-x-G-E.

NAME: Cullin family signature.

CONSENSUS: [LIV]-K-x(2)-[LIV]-x(2)-L-I-[DEQ]-[KRHNQ]-x-Y-[LIVM]-x-R-x(6,7)-[FY]-x-

CONSENSUS: Y-x-[SA]>.

NAME: Cullin family profile.

NAME: Enhancer of rudimentary signature.

CONSENSUS: Y-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S.

NAME: G10 protein signature 1.

CONSENSUS: L-C-C-x-[KR]-C-x(4)-[DE]-x-N-x(4)-C-x-C-R-V-P.

NAME: G10 protein signature 2.

CONSENSUS: C-x-H-C-G-C-[KRH]-G-C-[SA].

NAME: Glucokinase regulatory protein family signature.

CONSENSUS: G-[PA]-E-x-[LIV]-[STA]-G-S-[ST]-R-[LIVM]-K-[STGA](3)-x(2)-K.

NAME: GTP1/OBG family signature.

CONSENSUS: D-[LIVM]-P-G-[LIVM](2)-[DEY]-[GN]-A-x(2)-G-x-G.

NAME: HIT family signature.

CONSENSUS: [NQA]-x(4)-[GAV]-x-[QF]-x-[LIVM]-x-H-[LIVMFYT]-H-[LIVMFT]-H-[LIVMF](2)-

CONSENSUS: [PSGA].

NAME: Caseins alpha/beta signature.

CONSENSUS: C-L-[LV]-A-x-A-[LVF]-A.

NAME: Clathrin adaptor complexes medium chain signature 1.

CONSENSUS: [IVT]-[GSP]-W-R-x(2,3)-[GAD]-x(2)-[HY]-x(2)-N-x-[LIVMAFY](3)-D-[LIVM]-

CONSENSUS: [LIVMT]-E.

NAME: Clathrin adaptor complexes medium chain signature 2.

CONSENSUS: [LIV]-x-F-I-P-P-x-G-x-[LIVMFY]-x-L-x(2)-Y.

NAME: Clathrin adaptor complexes small chain signature.

CONSENSUS: [LIVM](2)-Y-[KR]-x(4)-L-Y-F.

NAME: Ependymins signature 1.

CONSENSUS: F-E-E-G-x-[LIVMF]-Y-[ED]-I-D-x(2)-N-[QE]-S-C-[RKH](2).

NAME: Ependymins signature 2.

CONSENSUS: [QE]-[LIVMA]-F-x(2)-P-[STA]-[FY]-C-[DE]-[GA]-[LIVM]-x(2)-[DE](2).

NAME: Syntaxin / epimorphin family signature.

CONSENSUS: [RQ]-x(3)-[LIVMA]-x(2)-[LIVM]-[ESH]-x(2)-[LIVMT]-x-[DEV]-[LIVM]-x(2)-

CONSENSUS: [LIVM]-[FS]-x(2)-[LIVM]-x(3)-[LIVT]-x(2)-Q-[GADEQ]-x(2)-[LIVM]-[DNQT]-x-

CONSENSUS: [LIVMF]-[DESV]-x(2)-[LIVM].

NAME: Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 1.

CONSENSUS: [GDER]-H-[FYWH]-T-Q-[LIVM](2)-W-x(2)-[STN].

NAME: Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2.

CONSENSUS: [LIVMFYH]-[LIVMFY]-x-C-[NQRHS]-Y-x-[PARH]-x-[GL]-N-[LIVMFYWDN].

NAME: Fetuin family signature 1.

CONSENSUS: C-x(56)-C-x(10)-C-x(13)-C-x(17,18)-C-x(13)-C-x(2)-C-x(58)-C-x(10,11)-

CONSENSUS: C-x(10,12)-C-x(16,22)-C.

NAME: Fetuin family signature 2.

CONSENSUS: L-E-T-x-C-H-x-L-D-P-T-P.

NAME: Legume lectins beta-chain signature.

CONSENSUS: [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST].

NAME: Legume lectins alpha-chain signature.

CONSENSUS: [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIV]-G-[LF]-[ST].

NAME: Vertebrate galactoside-binding lectin signature.

CONSENSUS: W-[GEK]-x-[EQ]-x-[KRE]-x(3,6)-[PCTF]-[LIVMF]-[NQEGRSKV]-x-[GH]-x(3)-

CONSENSUS: [DENKHS]-[LIVMFC].

NAME: Lysosome-associated membrane glycoproteins duplicated domain signature.

CONSENSUS: [STA]-C-[LIVM]-[LIVMFYW]-A-x-[LIVMFYW]-x(3)-[LIVMFYW]-x(3)-Y.

NAME: LAMP glycoproteins transmembrane and cytoplasmic domain signature.

CONSENSUS: C-x(2)-D-x(3,4)-[LIVM](2)-P-[LIVM]-x-[LIVM]-G-x(2)-[LIVM]-x-G-[LIVM](2)-

CONSENSUS: x-[LIVM](4)-A-[FY]-x-[LIVM]-x(2)-[KR]-[RH]-x(1,2)-[STAG](2)-Y-[EQ].

NAME: Glycophorin A signature.

CONSENSUS: I-I-x-[GAC]-V-M-A-G-[LIVM](2).

NAME: PMP-22 / EMP / MP20 family signature 1.

CONSENSUS: [LIVMF](4)-[SA]-T-x(2)-[DNKS]-x-W-x(9,13)-[LIV]-W-x(2)-C.

NAME: PMP-22 / EMP / MP20 family signature 2.

CONSENSUS: [RQ]-[AV]-x-M-[IV]-L-S-x-[LI]-x(4)-[GSA]-[LIVMF](3).

NAME: Oxysterol-binding protein family signature.

CONSENSUS: E-[KQ]-x-S-H-[HR]-P-P-x-[STACF]-A.

NAME: Yeast PIR proteins repeats signature.

CONSENSUS: S-Q-[IV]-[STGNH]-D-G-Q-[LIV]-Q-[AIV]-[STA].

NAME: Seminal vesicle protein I repeats signature.

CONSENSUS: [IVM]-x-G-Q-D-x-V-K-x(5)-[KN]-G-x(3)-[STLV].

NAME: Seminal vesicle protein II repeats signature.

CONSENSUS: [GSA]-Q-x-K-S-[FY]-x-Q-x-K-[SA].

NAME: Serum amyloid A proteins signature.

CONSENSUS: A-R-G-N-Y-[ED]-A-x-[QKR]-R-G-x-G-x-W-A.

NAME: Spermadhesins family signature 1.

CONSENSUS: C-G-x(2)-[LI]-x(4)-G-x-I-x(9)-C-x-W-T.

NAME: Spermadhesins family signature 2.

CONSENSUS: C-x-K-E-x-[LIVM]-E-[LIVM]-x-[DE]-x(3)-[GS]-x(5)-K-x-C.

NAME: Stress-induced proteins SRP1/TIP1 family signature.

CONSENSUS: P-W-Y-[ST](2)-R-L.

NAME: Glypicans signature.

CONSENSUS: C-x(2)-C-x-G-[LIVM]-x(4)-P-C-x(2)-[FY]-C-x(2)-[LIVM]-x(2)-G-C.

NAME: Syndecans signature.

CONSENSUS: [FY]-R-[IM]-[KR]-K(2)-D-E-G-S-Y.

NAME: Tissue factor signature.

CONSENSUS: W-K-x-K-C-x(2)-T-x-[DEN]-T-E-C-D-[LIVM]-T-D-E.

NAME: Translationally controlled tumor protein signature 1.

CONSENSUS: [IA]-G-[GAS]-N-[PA]-S-A-E-[GDE]-[PAGE]-x(0,1)-[DEG]-x-[DEN]-x(2)-[DE].

NAME: Translationally controlled tumor protein signature 2.

CONSENSUS: [FL]-[FY]-[TVT]-G-E-x-[MA]-x(2,5)-[DEN]-[GAS]-x-[LV]-[AV]-x(3)-[FY]-[KR]-

CONSENSUS: [DE].

NAME: Tub family signature 1.

CONSENSUS: F-[KHQ]-G-R-V-[ST]-x-A-S-V-K-N-F-Q.

NAME: Tub family signature 2.

CONSENSUS: A-F-[AG]-I-[SAC]-[LIVM]-[ST]-S-F-x-[GST]-K-x-A-C-E.

NAME: HCP repeats signature.

CONSENSUS: H-R-H-R-G-H-x(2)-[DE](7).

NAME: Bacterial ice-nucleation proteins octamer repeat.

CONSENSUS: A-G-Y-G-S-T-x-T.

NAME: Cell cycle proteins fisW / rodA / spoVE signature.

CONSENSUS: [NV]-x(5)-[GTR]-[LIVMA]-x-P-[PTLIVM]-x-G-[LIVM]-x(3)-[LIVMFW](2)-S-[YSA]-

CONSENSUS: G-G-[STN]-[SA].

NAME: Enterobacterial virulence outer membrane protein signature 1.

CONSENSUS: G-[LIVMFY]-N-[LIVM]-K-Y-R-Y-E.

NAME: Enterobacterial virulence outer membrane protein signature 2.

CONSENSUS: [FYW]-x(2)-G-x-G-Y-[KR]-F>.

NAME: Hydrogenases expression/synthesis hypA family signature.

CONSENSUS: F-[CSA]-[FY]-[DE]-[LIVA](2)-x(3)-[ST]-[LIVM]-x(16)-C-x(2)-C-x(12,15)-

CONSENSUS: C-P-x-C.

NAME: Hydrogenases expression/synthesis hupF/hupC family signature.

CONSENSUS: <M-C-[LIV]-[GA]-[LIV]-P-x-[QKR]-[LIV].

NAME: Staphylocoagulase repeat signature.

CONSENSUS: A-R-P-x(3)-K-x-S-x-T-N-A-Y-N-V-T-T-x(2)-[DN]-G-x(3)-Y-G.

NAME: 11-S plant seed storage proteins signature.

CONSENSUS: N-G-x-[DE](2)-x-[LIVMF]-C-[ST]-x(11,12)-[PAG]-D.

NAME: Dehydrins signature 1.

CONSENSUS: S(5)-[DE]-x-[DE]-G-x(1,2)-G-x(0,1)-[KR](4).

NAME: Dehydrins signature 2.

CONSENSUS: [KR]-[LIM]-K-[DE]-K-[LIM]-P-G.

NAME: Germin family signature.

CONSENSUS: G-x(4)-H-x-H-P-x-A-x-E-[LIVM].

NAME: Olcosins signature.

CONSENSUS: [AG]-[ST]-x(2)-[AG]-x(2)-[LIVM]-[SAD]-T-P-[LIVMF](4)-F-S-P-[LIVM](3)-

CONSENSUS: P-A.

NAME: Small hydrophilic plant seed proteins signature.

CONSENSUS: G-[EQ]-T-V-V-P-G-G-T.

NAME: Pathogenesis-related proteins Betv1 family signature.

CONSENSUS: G-x(2)-[LIVMF]-x(4)-E-x(2)-[CSTAEN]-x(8,9)-[GND]-G-[GS]-[CS]-x(2)-K-x(4)-

CONSENSUS: [FY].

NAME: Pollen proteins Ole e I family signature.

CONSENSUS: [EQ]-G-x-V-Y-C-D-T-C-R.

NAME: Thaumatin family signature.

CONSENSUS: G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-G-x(2,3)-C.

NAME: Mrp family signature.

CONSENSUS: W-x(2)-[LIVM]-D-[LIVMY](4)-D-x-P-P-G-T-[GS]-D.

NAME: Glucose inhibited division protein A family signature 1.

CONSENSUS: [GS]-P-x-Y-C-P-S-[LIVM]-E-x-K-[LIVM]-x-[KR]-F.

NAME: Glucose inhibited division protein A family signature 2.

CONSENSUS: A-G-Q-x-[NT]-G-x(2)-G-Y-x-E-[SAG](3)-[QS]-G-[LIVM](2)-A-G-[LIVMT]-N-A.

NAME: NOL1/NOP2/sun family signature.

CONSENSUS: [FV]-D-[KRA]-[LIVMA]-L-x-D-[AV]-P-C-[ST]-[GA].

NAME: PET112 family signature.

CONSENSUS: [DN]-x-[DN]-R-x(3)-P-L-[LIV]-E-[LIV]-x-[ST]-x-P.

NAME: Protein smpB signature.

CONSENSUS: [TA]-G-[LIVM]-x-L-x-G-x-E-[LIVM]-[KQ]-[SA]-[LIVM].

NAME: Hypothetical cof family signature 1.

CONSENSUS: [LIVFYAN]-[LIVMFA]-x(2)-D-[LIVMF]-[ND]-G-T-[LIV]-[LVY]-[STANLM].

NAME: Hypothetical cof family signature 2.
 CONSENSUS: [LIVMFC]-G-D-[GSANQ]-x-N-D-x(3)-[LIMFY]-x(2)-[AV]-x(2)-[GSCP]-x(2)-
 CONSENSUS: [LMP]-x(2)-[GAS].

NAME: RIO1/ZK632.3/MJ0444 family signature.
 CONSENSUS: [LIVM]-V-H-[GA]-D-L-S-E-[FY]-N-x-[LIVM].

NAME: SUA5/yciO/yrnC family signature.
 CONSENSUS: [LIVMTA](3)-[LIVMFYC]-[PG]-T-[DE]-[STA]-x-[FY]-[GA]-[LIVM]-[GS].

NAME: Uncharacterized protein family UPF0001 signature.
 CONSENSUS: [FW]-H-[FM]-[IV]-G-x-[LIV]-Q-x-[NKR]-K-x(3)-[LIV].

NAME: Uncharacterized protein family UPF0003 signature.
 CONSENSUS: G-x-V-x(2)-[LIV]-x(3)-[SA]-x(6)-D-x(3)-[LIVT](3)-P-N-x(2)-[LIVMF](2)-
 CONSENSUS: x(5)-N.

NAME: Uncharacterized protein family UPF0004 signature.
 CONSENSUS: [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVM]-x(4)-G.

NAME: Uncharacterized protein family UPF0005 signature.
 CONSENSUS: G-[LIVM](2)-[SA]-x(5,8)-G-x(2)-[LIVM]-G-P-x-L-x(4)-[SAG]-x(4,6)-
 CONSENSUS: [LIVM](2)-x(2)-A-x(3)-T-A-[LIVM](2)-F.

NAME: Uncharacterized protein family UPF0006 signature 1.
 CONSENSUS: [LIVMFY](2)-D-[STA]-H-x-H-[LIVMF]-[DN].

NAME: Uncharacterized protein family UPF0006 signature 2.
 CONSENSUS: P-[LIVM]-x-[LIVM]-H-x-R-x-[TA]-x-[DE].

NAME: Uncharacterized protein family UPF0006 signature 3.
 CONSENSUS: [LVSA]-[LIVA]-x(2)-[LIVM]-[PS]-x(3)-L-[LIVM]-[LIVMS]-E-T-D-x-P.

NAME: Uncharacterized protein family UPF0007 signature.
 CONSENSUS: V-L-[IV]-H-D-[GA]-A-R.

NAME: Uncharacterized protein family UPF0011 signature.
 CONSENSUS: S-D-A-G-x-P-x-[LIV]-[SN]-D-P-G.

NAME: Uncharacterized protein family UPF0012 signature.
 CONSENSUS: [GTA]-x(2)-[IVT]-C-Y-D-[LIVM]-x-F-P-x(9)-G.

NAME: Uncharacterized protein family UPF0015 signature.
 CONSENSUS: [DE]-[LIVMF](3)-R-T-[SG]-G-x(2)-R-x-S-x-[FY]-[LIVM](2)-W-Q.

NAME: Uncharacterized protein family UPF0016 signature.
 CONSENSUS: E-[LIVM]-G-D-K-T-F-[LIVMF](2)-A.

NAME: Uncharacterized protein family UPF0017 signature.
 CONSENSUS: D-x(8)-[GN]-[LFY]-x(4)-[DET]-[LY]-Y-x(3)-[ST]-x(7)-[IV]-x(2)-[PS]-x-
 CONSENSUS: [LIVM]-x-[LIVM]-x(3)-[DN]-D.

NAME: Uncharacterized protein family UPF0019 signature.
 CONSENSUS: L-P-V-[VT]-[NQL]-F-[AT]-A-G-G-[LIV]-A-T-P-A-D-A-A-[LM].

NAME: Uncharacterized protein family UPF0020 signature.
 CONSENSUS: D-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E.

NAME: Uncharacterized protein family UPF0021 signature.
 CONSENSUS: C-K-x(2)-F-x(4)-E-x(22,23)-S-G-G-K-D.

NAME: Uncharacterized protein family UPF0023 signature.
 CONSENSUS: D-x-D-E-[LIV]-L-x(4)-V-F-x(3)-S-K-G.

NAME: Uncharacterized protein family UPF0024 signature.
 CONSENSUS: G-x-K-D-[KR]-x-A-[LV]-T-x-Q-x-[LIVF]-[SGC].

NAME: Uncharacterized protein family UPF0025 signature.
 CONSENSUS: D-V-[LIV]-x(2)-G-H-[ST]-H-x(12)-[LIVMF]-N-P-G.

NAME: Uncharacterized protein family UPF0027 signature.
 CONSENSUS: Q-[LIVM]-x-N-x-A-x-[LIVM]-P-x-I-x(6)-[LIVM]-P-D-x-H-x-G-x-G-x(2)-[IV]-G.

NAME: Uncharacterized protein family UPF0028 signature.

CONSENSUS: [GA]-[GS]-G-[GA]-A-R-G-x-[SA]-H-x-G-x(9)-[IV]-x-[IV]-D-x(2)-[GA]-G-x-S-
 CONSENSUS: x-G.

NAME: Uncharacterized protein family UPF0029 signature.

CONSENSUS: G-x(2)-[LIVM](2)-x(2)-[LIVM]-x(4)-[LIVM]-x(5)-[LIVM](2)-x-R-[FYW](2)-G-
 CONSENSUS: G-x(2)-[LIVM]-G.

NAME: Uncharacterized protein family UPF0030 signature.

CONSENSUS: [GA]-L-I-[LIV]-P-G-G-E-S-T-[STA].

NAME: Uncharacterized protein family UPF0031 signature 1.

CONSENSUS: [SAV]-[IVW]-[LVA]-[LIV]-G-[PNS]-G-L-[GP]-x-[DENQT].

NAME: Uncharacterized protein family UPF0031 signature 2.

CONSENSUS: [GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM].

NAME: Uncharacterized protein family UPF0032 signature.

CONSENSUS: Y-x(2)-F-[LIVMA](2)-x-L-x(4)-G-x(2)-F-[EQ]-[LIVMF]-P-[LIVM].

NAME: Uncharacterized protein family UPF0033 signature.

CONSENSUS: L-[DN]-x(2)-[TAG]-x(2)-C-P-x-P-x-[LIVM].

NAME: Uncharacterized protein family UPF0034 signature.

CONSENSUS: [LIVM]-[DNG]-[LIVM]-N-x-G-C-P-x(3)-[LIVMASQ]-x(5)-G-[SAC].

NAME: Uncharacterized protein family UPF0035 signature.

CONSENSUS: L-L-T-x-R-[SA]-x(3)-R-x(3)-G-x(3)-F-P-G-G.

NAME: Uncharacterized protein family UPF0036 signature.

CONSENSUS: H-x-S-G-H-[GA]-x(3)-[DE]-x(3)-[LM]-x(5)-P-x(3)-[LIVM]-P-x-H-G-[DE].

NAME: Uncharacterized protein family UPF0038 signature.

CONSENSUS: G-x-[LI]-x-R-x(2)-L-x(4)-F-x(8)-[LIV]-x(5)-P-x-[LIV].

NAME: Uncharacterized protein family UPF0044 signature.

CONSENSUS: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-
 CONSENSUS: x(2)-G.

NAME: Uncharacterized protein family UPF0047 signature.

CONSENSUS: S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV].

NAME: Uncharacterized protein family UPF0054 signature.

CONSENSUS: H-[GS]-x-L-H-L-[LI]-G-[FYW]-D-H.

NAME: Uncharacterized protein family UPF0057 signature.

CONSENSUS: [LIV]-x-[STA]-[LIVF](3)-P-P-[LIVA]-[GA]-[IV]-x(4)-[GKN].

NAME: Hypothetical YER057c/yjjV family signature.

CONSENSUS: P-[AT]-R-[SA]-x-[LIVMY]-x(2)-[AK]-x-L-P-x(4)-[LIVM]-E.

NAME: Hypothetical hesB/yadR/yfhF family signature.

CONSENSUS: F-x-[LIVMFY]-x-N-[PG]-[NSK]-x(4)-C-x-C-[GS]-x-S-F.

NAME: Hypothetical yabO/yceC/sfhB family signature.

CONSENSUS: [NHY]-R-[LI]-D-x(2)-T-[ST]-G-[LIVMA]-[LIVMF](2)-[LIVMFG]-[SGAC].

We claim:

1. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3b16; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_4l1m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_7l1o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; htes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06;

htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15;
 htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9;
 htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20;
 Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1;
 hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2;
 hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15;
 hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11;
 hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

2. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4; hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10; hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; their complements; and variants thereof.

3. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16f21; hfbr2_16k22; hfbr2_22f21; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23f2; ; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2c1; hfbr2_2c18; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2k19; hfbr2_3f16; hfbr2_3l2; hfbr2_62n10; hfbr2_64a11; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64o16; hfbr2_6a17; hfbr2_6i20; hfbr2_71o20;

hfbr2_72d13; hfbr2_72m16; hfbr2_72n12; hfbr2_78d13; hfbr2_78n23; hfbr2_7a24;
hfbr2_7e22; hfbr2_7j4; hfbr2_82m16; and hfbr1_10.

4. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; their complements; and variants thereof.

5. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24e23; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_4b6; hfkd2_4c8; their complements; and variants thereof.

6. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; their complements; and variants thereof.

7. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1c23 hmcfl_1g13; their complements; and variants thereof.

8. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17i17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18i7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21i16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23i11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2i19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21;

htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9;
htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19;
htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11;
Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10;
htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22;
Htes3_9i20; Htes3_9k22; their complements; and variants thereof.

9. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: htes3_14g5; htes3_14p14; htes3_14p7; htes3_15a13; htes3_15g14; htes3_15h1; htes3_15j18; htes3_17f10; Htes3_18f3; htes3_19f19; htes3_19j17; htes3_20c21; htes3_21n23; htes3_22c23; htes3_22n13; Htes3_23n19; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2d15; htes3_2f14; htes3_2g7; htes3_2h15; htes3_2i19; htes3_2m20; htes3_2n9; htes3_30f4; htes3_35g6; htes3_35n24; htes3_35p17; htes3_4b4; htes3_4f17; htes3_4o19; htes3_50j4; htes3_50n23; htes3_50n06; htes3_6b21; htes3_6d16; htes3_72k11; htes3_7d17; htes3_7j8; Htes3_8g11; Htes3_8g5; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; their complements; and variants thereof.

10. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16g18; hfbr2_2k14; Htes3_35b4; htes3_35p22; htes3_7j3; htes3_7p10; hute1_20m11; their complements; and variants thereof.

11. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_2b5; htes3_15i5; htes3_18i7; htes3_1k11; Htes3_72k15; htes3_7b22; hute1_19g22; hute1_24j6; their complements; and variants thereof.

12. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_2d15; htes3_35e21; hute1_2h3; their complements; and variants thereof.

13. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23i24; hfbr2_2i17; hfbr2_41m15; hfbr2_62f10; hfbr2_62i19; hfbr2_64j18;

hfkd2_24n20; hfkd2_24p5; hfkd2_4k14; htes3_1g13; htes3_21l16; htes3_23l11; htes3_26g22; htes3_4h6; htes3_72p16; hute1_19h17; hute1_20h13; hute1_24e11; their complements; and variants thereof.

14. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_3g8; hfbr2_62o17; hfbr2_6b24; hfbr2_78k24; hfkd2_24b15; hfkd2_3o17; hfkd2_46j20; htes3_17l17; htes3_17n18; htes3_27d1; htes3_2a17; htes3_35b5; htes3_35k16; htes3_35n12; htes3_35n9; hute1_20b19; hute1_20m24; hute1_23e13; their complements; and variants thereof.

15. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23b10; hfbr2_3c18; hfbr2_64a15; hfbr2_6o17; hfbr2_72b18; hfbr2_72l12; hfbr2_82i24(hfbr1_10); htes3_14h21; Htes3_15j3; htes3_20m18; htes3_22g2; htes3_2m18; htes3_7p9; htes3_8m10; hute1_18l1; their complements; and variants thereof.

16. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23b21; hfbr2_23n16; hfbr2_2c17; hfbr2_62b11; hfbr2_78c24; hfbr2_82e4 (hfbr1_10e4); hfbr2_82i17 (hfbr1_10); hfbr2_82m6 (hfbr1_10); hfkd2_46m4; htes3_15k11; htes3_1c1; htes3_1n3; htes3_20k2; htes3_21d4; htes3_23n19; htes3_4f5; htes3_6c11; htes3_8e24; hute1_20g21; hute1_22d2; hute1_22e12; their complements; and variants thereof.

17. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16i12; hfbr2_16l12; hfbr2_22h13; hfbr2_2b17; hfbr2_2d17; hfbr2_64k24; hfbr2_82c20 (hfbr1_10c20); hfbr2_82e17 (hfbr1_10e17); hfbr2_82g14 (hfbr1_10g14); hfkd2_24a15; hfkd2_3i13; hfkd2_4m11; hmcfl_1a11; hmcfl_1e15; htes3_15c6; htes3_2o13; htes3_27k4; htes3_2h1; htes3_35k24; hute1_19f19; and hute1_24c19; their complements; and variants thereof.

18. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_46k19; hfkd2_47a4;

htes3_2e12; htes3_21j15; htes3_17n12; hute1_18i19; hute1_1i2; their complements; and variants thereof.

19. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1; hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

20. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hute1_17k7; hute1_18c12; hute1_18i4; hute1_19g19; hute1_19j11; hute1_22n2; hute1_21d15; hute1_22o2; hute1_23g11; their complements; and variants thereof.

21. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6;

hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4;
 hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15;
 hmcfl_1g13; hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7;
 htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5;
 htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17i17; htes3_17n12;
 htes3_17n18; Htes3_18f3; htes3_18i7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13;
 htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15;
 htes3_21i16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23i11;
 htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14;
 htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7;
 htes3_2h1; htes3_2h15; htes3_2i19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13;
 htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16;
 htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22;
 htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06;
 htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15;
 htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9;
 htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20;
 Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18i1;
 hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2;
 hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15;
 hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11;
 hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants
 thereof.

22. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12;
 hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8;
 hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23i24; ; hfbr2_23n16; hfbr2_23o24;
 hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18;
 hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17;
 hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3i2; hfbr2_41m15;
 hfbr2_62b11; hfbr2_62f10; hfbr2_62i19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11;

hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18;
hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17;
hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12;
hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22;
hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;
hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10;
hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; complements of the nucleic acid
sequences; and variants thereof.

23. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfbr2_16f21; hfbr2_16k22; hfbr2_22f21; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8;
hfbr2_23f2; ; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2c1; hfbr2_2c18; hfbr2_2d20;
hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2k19; hfbr2_3f16; hfbr2_3l2; hfbr2_62n10;
hfbr2_64a11; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64k24;
hfbr2_64o16; hfbr2_6a17; hfbr2_6i20; hfbr2_71o20; hfbr2_72d13; hfbr2_72m16;
hfbr2_72n12; hfbr2_78d13; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82m16;
hfbr1_10; complements of the nucleic acid sequences; and variants thereof.

24. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5;
hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20;
hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14;
hfkd2_4m11; complements of the nucleic acid sequences; and variants thereof.

25. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of: hfkd2_1j9;
hfkd2_24e23; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_4b6; hfkd2_4c8;
complements of the nucleic acid sequences; and variants thereof.

26. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:

hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

27. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmcfl_1c23; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

28. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17i17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18i7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21i16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23i11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2i19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; complements of the nucleic acid sequences; and variants thereof.

29. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: htes3_14g5; htes3_14p14; htes3_14p7; htes3_15a13; htes3_15g14; htes3_15h1; htes3_15j18; htes3_17f10; htes3_17n18; Htes3_18f3; htes3_19f19; htes3_19j17; htes3_20c21; htes3_21n23; htes3_22c23; htes3_22n13; Htes3_23n19; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2d15; htes3_2f14; htes3_2g7; htes3_2h15; htes3_2i19; htes3_2m20; htes3_2n9; htes3_30f4; htes3_35g6; htes3_35n24; htes3_35p17; htes3_4b4; htes3_4f17;

htes3_4o19; htes3_50j4; htes3_50n23; htes3_50n06; htes3_6b21; htes3_6d16; htes3_72k11; htes3_7d17; htes3_7j8; Htes3_8g11; Htes3_8g5; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; complements of the nucleic acid sequences; and variants thereof.

30. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16g18; hfbr2_2k14; Htes3_35b4; htes3_35p22; htes3_7j3; htes3_7p10; hute1_20m11; complements of the nucleic acid sequences; and variants thereof.

31. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_2b5; htes3_15i5; htes3_18l7; htes3_1k11; Htes3_72k15; htes3_7b22; hute1_19g22; hute1_24j6; complements of the nucleic acid sequences; and variants thereof.

32. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_2d15; htes3_35e21; hute1_2h3; complements of the nucleic acid sequences; and variants thereof.

33. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_23l24; hfbr2_2i17; hfbr2_41m15; hfbr2_62f10; hfbr2_62l19; hfbr2_64j18; hfkd2_24n20; hfkd2_24p5; hfkd2_4k14; htes3_1g13; htes3_21l16; htes3_23l11; htes3_26g22; htes3_4h6; htes3_72p16; hute1_19h17; hute1_20h13; hute1_24e11; complements of the nucleic acid sequences; and variants thereof.

34. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_3g8; hfbr2_62o17; hfbr2_6b24; hfbr2_78k24; hfkd2_24b15; hfkd2_3o17; hfkd2_46j20; htes3_17l17; Htes3_17n18; htes3_27d1; htes3_2a17; htes3_35b5; htes3_35k16; htes3_35n12; htes3_35n9; hute1_20b19; hute1_20m24; hute1_23e13; complements of the nucleic acid sequences; and variants thereof.

35. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfbr2_23b10; hfbr2_3c18; hfbr2_64a15; hfbr2_6o17; hfbr2_72b18; hfbr2_72l12;
hfbr2_82i24(hfbr1_10); htes3_14h21; Htes3_15j3; htes3_20m18; htes3_22g2; htes3_2m18;
htes3_7p9; htes3_8m10; hute1_18l1; complements of the nucleic acid sequences; and
variants thereof.

36. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfbr2_23b21; hfbr2_23n16; hfbr2_2c17; hfbr2_62b11; hfbr2_78c24; hfbr2_82e4
(hfbr1_10e4); hfbr2_82i17 (hfbr1_10); hfbr2_82m6 (hfbr1_10); hfkd2_46m4; htes3_15k11;
htes3_1c1; hhtes3_1n3; htes3_20k2; htes3_21d4; htes3_23n19; htes3_4f5; htes3_6c11;
htes3_8e24; hute1_20g21; hute1_22d2; hute1_22e12; complements of the nucleic acid
sequences; and variants thereof.

37. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfbr2_16i12; hfbr2_16l12; hfbr2_22h13; hfbr2_2b17; hfbr2_2d17; hfbr2_64k24;
hfbr2_82c20 (hfbr1_10c20); hfbr2_82e17 (hfbr1_10e17); hfbr2_82g14 (hfbr1_10g14);
hfkd2_24a15; hfkd2_3i13; hfkd2_4m11; hmcfl_1a11; hmcfl_1e15; htes3_15c6;
htes3_2ol3; htes3_27k4; htes3_2h1; htes3_35k24; hute1_19f19; and hute1_24c19;
complements of the nucleic acid sequences; and variants thereof.

38. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfkd2_46k19; hfkd2_47a4; htes3_2e12; htes3_21j15; htes3_17n12; hute1_18i19;
hute1_1i2; complements of the nucleic acid sequences; and variants thereof.

39. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1; hute1_19f19;
hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19;
hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2;
hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19;
hute1_24e11; hute1_24j6; hute1_2h3; complements of the nucleic acid sequences; and
variants thereof.

40. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hute1_17k7; hute1_18c12; hute1_18i4; hute1_19g19; hute1_19j11; hute1_22n2; hute1_21d15; hute1_22o2; hute1_23g11; complements of the nucleic acid sequences; and variants thereof.

41. A nucleic acid molecule having the sequence of a clone selected from the group consisting of hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3b16; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4; hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10; hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; htes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18;

htes3_2m20; htes3_2n9; htes3_2ol3; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1; hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

42. A polypeptide encoded by the nucleic acid molecule according to claim 41.

43. An antibody or fragment thereof that is capable of binding to a specific portion of the peptide according to claim 42.

44. A pharmaceutical composition, comprising (a) an effective amount of a pharmaceutical agent, wherein said pharmaceutical agent is selected from the group consisting of the polypeptide according to claim 42, variants or functional derivatives thereof, and antibodies thereto; and (2) a physiologically acceptable carrier or excipient.

45. An expression vector comprising the nucleic acid molecule of claim 41 or a fragment thereof, and optionally a promoter operably linked to said nucleic acid molecule or said fragment.

46. A method for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by the nucleic acid molecule according to claim 41.

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A3

(51) International Patent Classification⁷: **C12N 15/12**,
C07K 14/47, C12Q 1/68, C07K 16/18, A61K 38/17, C12P
21/00

(21) International Application Number: PCT/IB00/01496

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)

(71) Applicant (for all designated States except US): **FRAUN-
HOFER-GESELLSCHAFT ZUR FOERDERUNG
DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE];
Leonrodstrasse 54, D-80636 München (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WIEMANN, Stefan** [DE/DE]; Grosse Lachstrasse 30a, 69207 Sandhausen (DE). **POUSTKA, Annemarie** [DE/DE]; Ladenburgstrasse 41, 69120 Heidelberg (DE). **WELLENREUTHER, Ruth** [DE/DE]; An der Markscheide 5, 69126 Heidelberg (DE). **BLUM, Helmut** [DE/DE]; Koenigswieser Strasse 94, 81475 Muenchen (DE). **OBERRMAIER, Brigitte** [DE/DE]; Muehlstrasse 9a, 82547 Eurasberg (DE). **OTTENWAEELDER, Birgit** [DE/DE]; Beinhofstrasse 1a, 81247 Muenchen (DE). **BAHR, André** [DE/DE]; Ratfaelweg 6, 40724 Hilden (DE). **DUESTERHOEFT, Andreas** [DE/DE]; Karlrobert-Kreiten-Strasse 14, 40724 Hilden (DE). **KOENIG, Christoph** [DE/US]; 6233 22nd Avenue N.E., Seattle, WA 98115 (US). **LAUBER, Juergen** [DE/DE]; Unterberg 1F, 42799 Leichlingen (DE). **HEUBNER, Dagmar** [DE/DE]; Gruene Trift 126a, 12557 Berlin (DE). **WAMBUTT, Rolf** [DE/DE]; Florian-Geyer-Strasse 28, 12489 Berlin (DE). **KOEHRER, Karl** [DE/DE]; Schlossmannstrasse 4, 40225 Duesseldorf (DE). **BEYER, Andreas** [DE/DE];

Helgolandring 106, 45149 Essen (DE). **GASSENHUBER, Johann** [DE/DE]; Emanuel Geibel Strasse 8, 65185 Wiesbaden (DE). **GRUBER, Christian** [DE/DE]; Zasinger Strasse 8, 81547 Muenchen (DE). **STRACK, Norman** [DE/DE]; Linderbergweg 1, 82229 Seefeld (DE). **MEWES, H.W.** [DE/DE]; Graf Toerring Strasse 9, 82237 Woerthsee (DE). **ANSORGE, Wilhelm** [DE/DE]; Boxberring 107/55, 69126 Heidelberg (DE). **GLASSL, Sabine** [DE/DE]; Friedberger Weg 2, 64720 Michelstadt (DE). **RITTMUELLER, Claudia** [DE/DE]; Siedlerweg 2, 69151 Dilsbergerhof (DE). **REGIERT, Thomas** [DE/DE]; Raiffeisenstrasse 38, 67227 Frankenthal (DE). **BLOECKER, Helmut** [DE/DE]; Doeringstrasse 16, 38118 Braunschweig (DE). **BOECHER, Michael** [DE/DE]; Alter Weg 41a, 38302 Wolfenbuettel (DE). **HORNISCHER, Klaus** [DE/DE]; Mozartstrasse 2, 38106 Braunschweig (DE). **NORDSIEK, Gabriele** [DE/DE]; Ohfeld 34, 31188 Holle (DE). **TAMPE, Jens** [DE/DE]; Bergisch-Gladbacher-Strasse 656, 51067 Koeln (DE).

(74) Agents: **MERCER, Christopher, Paul et al.**; Carpmals & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

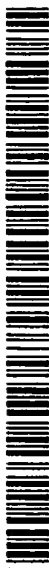
— with international search report

(88) Date of publication of the international search report:
20 June 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.



WO 01/12659 A3

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01496

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C12Q1/68 C07K16/18 A61K38/17
C12P21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 09552 A (GENETICS INST) 24 February 2000 (2000-02-24) Page 546, claim 86: SEQ.ID.No.: 77 ---	1-46
X	HILLIER L ET AL: "Human cDNA clone IMAGE:754267" EMBL SEQUENCE DATABASE, 23 July 1997 (1997-07-23), XP002163418 HEIDELBERG DE Accession Nr.: AA478899 abstract --- -/--	1-42

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
 "E" earlier document but published on or after the international filing date
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 "O" document referring to an oral disclosure, use, exhibition or other means
 "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
 "3" document member of the same patent family

Date of the actual completion of the international search

20 March 2001

Date of mailing of the international search report

07.06.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

De Kok, A

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01496

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HILLIER L ET AL.: "Human cDNA clone IMAGE: 754167" EMBL SEQUENCE DATABASE, 23 June 1997 (1997-06-23), XP002163419 HEIDELBERG DE Accession Nr.: AA478780 abstract ---	1-42
X	STRAUSBERG R ET AL.: "Human cDNA sequence IMAGE:2138166" EMBL SEQUENCE DATABASE, 24 March 1999 (1999-03-24), XP002163420 HEIDELBERG DE Accession Nr.:522149 abstract ---	1-42
X	HILLIER L ET AL.: "Human cDNA clone IMAGE:263887" EMBL SEQUENCE DATABASE, 5 January 1996 (1996-01-05), XP002163421 HEIDELBERG DE Accession Nr.: N28525 abstract ---	1-42
A	"Atlas(tm) human cDNA expression array I" CLONTECHNIQUES, April 1977 (1977-04), pages 4-7, XP002914393 US the whole document ---	1-20
A	REICHERT J ET AL: "HUMAN AND RODENT EXPRESSION PATTERN OF A FUSION GENE ISOLATED FROM AN MCF7 CDNA LIBRARY" INTERNATIONAL JOURNAL OF ONCOLOGY, vol. 9, no. 1, 1996, pages 29-32, XP000906725 page 29 ---	1,6,7,17
A	WO 98 40486 A (GENETICS INST) 17 September 1998 (1998-09-17) page 29, line 20 -page 60, line 13 page 18, line 5 -page 26, line 32 -----	1-5, 8-25, 28-46

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 00/01496

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 21-40
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(v) PCT - Presentation of information:
Although claims 21-40 could be considered as a mere presentation of information, according to Rule 39.1(v) PCT, the search has been carried out as far as possible in our systematic documentation.
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-46 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-46, all partially

Invention 1:

A nucleic acid molecule having the sequence of the clone hfbr2_16c16 (corresponding to SEQ.ID.1); an assemblage comprising said nucleic acid; a computer readable medium comprising said nucleic acid; a polypeptide encoded by said nucleic acid; an antibody binding to said polypeptide; an expression vector comprising said nucleic acid and a method for producing said polypeptide.

2. Claims: 1-46, all partially

Invention 2-233:

same as invention 1, but for each single clone as set forth in claim 1 (i.e. starting with clone hfbr2_16f21 and ending with clone hutel_2h3)

NB: for the sake of conciseness, the first subject-matter is explicitly defined, the other subject-matter by analogy thereto.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 00/01496

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0009552 A	24-02-2000	AU 5557099 A	06-03-2000
WO 9840486 A	17-09-1998	US 5976837 A	02-11-1999
		AU 6702298 A	29-09-1998
		EP 0973890 A	26-01-2000